An Integrative Genomic Analysis of the Relationship between Ageing, Cancer, and Development

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

Kasit Chatsirisupachai

July 2022

Abstract

Ageing can be broadly described as a progressive decline in physiological functions, leading to increased susceptibility to diseases and mortality. Ageing is the most important risk factor for various diseases, including cardiovascular diseases, neurodegenerative diseases, and most types of cancer. However, how ageing processes influence cancer development remains largely unclear. Furthermore, how molecular alterations in tumours differ among patients of different ages has not been systematically explored. In addition, previous studies have suggested that some aspects of ageing may be partly originated from developmental processes. Yet, how the molecular changes during prenatal development relate to those changes during ageing remains to be elucidated. In this thesis, I leveraged multiple large-scale genomic, transcriptomic, and epigenetic datasets that are publicly available to shed light on the relationship between ageing and cancer.

In chapter 2, I investigated the transcriptomic relationship between human ageing and cancer in a tissue-specific manner, using gene expression data from The Genotype-Tissue Expression (GTEx) and The Cancer Genome Atlas (TCGA). I found that ageing and cancer genes changed in the opposite direction in most tissues except thyroid and uterus. I further derived cellular senescence signatures from a meta-analysis and showed that senescence genes changed in the same direction as ageing genes and the opposite direction of cancer genes.

Next, I took advantage of the pan-cancer multi-omic cancer datasets from TCGA to systematically explore the age-associated molecular patterns across cancers in chapter 3. I showed that, in general, somatic copy-number alterations (SCNAs) and somatic mutations increased with age. Importantly, I identified age-associated SCNAs and somatic mutations in several cancer-driver genes across different cancer types. The largest age-related genomic differences were found in gliomas and endometrial cancer. I also identified age-related gene expression changes in cancer and demonstrated that these genes might be, in part, regulated by age-associated DNA methylation changes. The analysis in this chapter provided a comprehensive view of age-associated molecular alterations in cancer.

Chapter 4 studied the relationship between dynamic gene expression and histone modification changes during prenatal development and ageing in mice. I observed a significant overlap between development and ageing differentially expressed genes in all tissues tested. I classified differentially expressed genes by development and ageing trajectories and examined the biological functions of genes in each trajectory. Further investigation of the development and ageing transcriptomes also supported previous reports that transcriptomes across tissues diverge during development and tend to converge during ageing.

In summary, this thesis demonstrated the application of genomics and bioinformatics to better understand the links between ageing, cancer, and development. Results from this study shed light on how ageing influences cancer, underscore age as an important factor in cancer genotypes, and suggest potential links between mammalian development and ageing.

Acknowledgements

First of all, I would like to sincerely thank my primary supervisor, Prof João Pedro de Magalhães, for all your support, thoughts, encouragement, and kindness. I am very grateful to have a chance to work with you. I would also like to thank my secondary supervisor, Prof Luminita Paraoan, for your guidance and support. I would like to acknowledge Dr Peter Van Loo and Dr Tom Lesluyes, who are wonderful collaborators. I have learned a lot from working with you. I also appreciate the advice and support from my internal assessors, Dr Kazuhiro Yamamoto and Dr Rachel Oldershaw. I would also like to thank amazing friends and colleagues at the Integrative Genomics of Ageing Group. They have been very generous and supportive throughout my years in Liverpool.

I really appreciate continuous support from my advisors, Assoc Prof Arthit Chairoungdua and Assoc Prof Varodom Charoensawan from Faculty of Science, Mahidol University. Many thanks for always believing in me and for your encouragement. I am truly fortunate to be your student. I would also like to thank Mahidol-Liverpool PhD scholarship for the opportunity to come and study in Liverpool.

My special thanks go to my PhD crew, a group of friends from high school who are currently pursuing their PhD at various places around the world! Indeed, I would like to thank all Liverpool-Thai PhD student members for being such wonderful friends and family here. I will miss our party times for sure!

Last but not least, this thesis could not have happened without the unconditional love and supports from my parents and brother. Thanks for always being there for me and cheering me up. Finally, I would like to thank Dowrung, my best partner, personal cheerleader, and chef! Thank you for always believing in me. And I will be your number one supporter as well. You are the best!

Table of Contents

Table of Figures

Table of Tables

Abbreviations

AACR GENIE	American Association for Cancer Research: Genomics Evidence Neoplasia Information Exchange
ACC	adrenocortical carcinoma
age-DEGs	differentially expressed genes with age
age-DMGs	differentially methylated genes with age
ARDs	age-related diseases
BC	betweenness centrality
BLCA	bladder urothelial carcinoma
BRCA	breast invasive carcinoma
cancer-DEGs	differentially expressed genes in cancer
CC	closeness centrality
ccRCC	clear cell renal cell carcinoma
CESC	cervical squamous cell carcinoma and endocervical adenocarcinoma
CH	clonal haematopoiesis
CHIP	clonal haematopoiesis of indeterminate potential
ChIP-seq	chromatin immunoprecipitation sequencing
CHOL	cholangiocarcinoma
CI	confidence interval
COAD	colon adenocarcinoma
COSMIC	Catalogue of Somatic Mutations in Cancer
CoV	coefficient of variation
CPM	counts per million mapped reads
DDR	DNA damage response
DiCo	divergence followed by convergence
DLBC	lymphoid neoplasm diffuse large B-cell lymphoma
ECM	extracellular matrix
EMT	epithelial-to-mesenchymal transition
ENCODE	the Encyclopedia of DNA Element
ER	oestrogen receptor
ESCA	oesophageal carcinoma
FC	fold change

GBM	glioblastoma multiforme
GEO	Gene Expression Omnibus
GI	genomic instability
GO	gene ontology
GSEA	gene set enrichment analysis
GTEx	the Genotype Tissue Expression
H3K27ac	acetylation of histone 3 at lysine 27
H3K4me1	monomethylation of histone 3 at lysine 4
H3K4me3	trimethylation of histone 3 at lysine 4
H3K9me3	trimethylation of histone 3 at lysine 9
HER2	human epidermal growth factor receptor 2
HNSC	head and neck squamous cell carcinoma
HSCs	haematopoietic stem cells
ICGC	International Cancer Genome Consortium
IDH	isocitrate dehydrogenase
IIS	insulin/IGF-1 signalling
IL-6	interleukin-6
Indels	small insertions and deletions
IntOGen	Integrative OncoGenomics
KEGG	Kyoto Encyclopedia of Genes and Genomes
KICH	kidney chromophobe
KIRC	kidney renal clear cell carcinoma
KIRP	kidney renal papillary cell carcinoma
LAML	acute myeloid leukaemia
LGG	brain lower grade glioma
LIHC	liver hepatocellular carcinoma
LOH	loss of heterozygosity
LUAD	lung adenocarcinoma
LUSC	lung squamous cell carcinoma
MESO	mesothelioma
MMP	matrix metalloproteinases
MSI	microsatellite instability
mTOR	mechanistic target of rapamycin
OIS	oncogene-induced senescence

OR	odds ratio
OV	ovarian serous cystadenocarcinoma
PAAD	pancreatic adenocarcinoma
PCA	principal component analysis
PCAWG	Pan-Cancer Analysis of Whole Genome
PCPG	pheochromocytoma and paraganglioma
POLD1	polymerase δ
POLE	polymerase ε
PRAD	prostate adenocarcinoma
READ	rectum adenocarcinoma
RNA-seq	RNA sequencing
ROS	reactive oxygen species
RPKM	reads per kilobase of transcript per million mapped reads
SA-βGal	senescence-associated β-galactosidase
SAHFs	senescence-associated heterochromatic foci
SARC	sarcoma
SASPs	senescence-associated secretory phenotypes
SCNAs	somatic copy-number alterations
scRNA-seq	single-cell RNA sequencing
SKCM	skin cutaneous melanoma
SNPs	single-nucleotide polymorphisms
SNVs	single-nucleotide variants
STAD	stomach adenocarcinoma
SVs	structural variations
TCGA	The Cancer Genome Atlas
TGCT	testicular germ cell tumours
THCA	thyroid carcinoma
THYM	thymoma
TMM	trimmed mean of M values
TNF	tumour necrosis factor
UCEC	uterine corpus endometrial carcinoma
UCS	uterine carcinosarcoma
UVM	uveal melanoma
VAF	variant allele frequency

VEGF	vascular endothelial growth factor
WES	whole-exome sequencing
WGD	whole-genome duplication
WGS	whole-genome sequencing

Chapter 1

Introduction

1.1 The biology of ageing and cancer

1.1.1 Ageing and age-related diseases: A brief introduction

Ageing can be broadly described as a complex, progressive process associated with a decline in physiological function, leading to increased susceptibility to diseases and increased mortality (Kirkwood, 2005; de Magalhães & Stuart-Hamilton, 2011; Rose *et al.*, 2012; Lopez-Otin *et al.*, 2013). Ageing is associated with gradual and consistent changes in various phenotypes. The loss of muscle mass (sarcopenia), cognitive ageing, a decline in immune system function, and hair greying are constantly observed across a population. However, the biological mechanisms underlying these phenotypes are largely unclear. Over the years, scientists have been trying to explain ageing, and as a result, numerous theories of ageing have emerged. Two major classes of theories of ageing are 1) damage-based theories of ageing and 2) programmed theories of ageing (Jin, 2010). The damage-based theories of ageing argue that ageing is a consequence of damage accumulation over time, such as DNA damage and somatic mutations, damage caused by free radicals, and protein defects. These various types of damage result in cellular dysfunction, cell losses, disruption of tissue homeostasis, and eventually ageing. The programmed theories of ageing, on the other hand, propose that ageing occurs because of a non-stochastic genetically defined, evolutionarily conserved programme. More recently scientists attempted to identify and classify biological mechanisms associated with ageing, such as "the seven pillars of ageing" (**Figure 1.1 left**) (Kennedy *et al.*, 2014) and "the nine hallmarks of ageing" (**Figure 1.1 right**) (Lopez-Otin *et al.*, 2013). Even though these ageing concepts have evolved over time, the central paradigm to explain ageing is still lacking, positing the challenge to understand mechanistic causes of ageing (Gems & de Magalhaes, 2021).

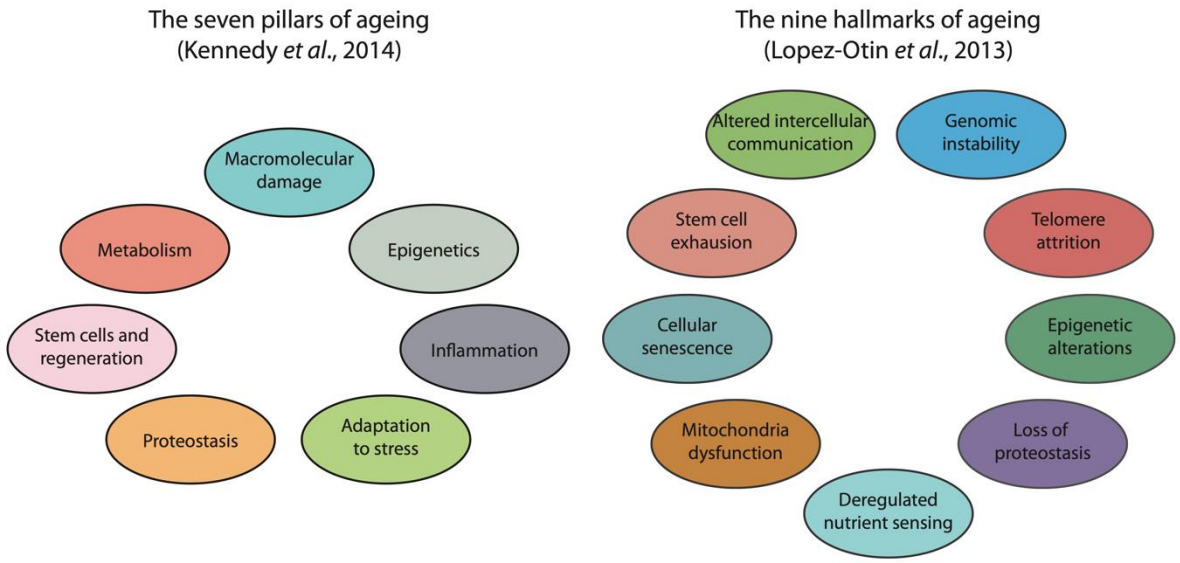


Figure 1.1. Biological mechanisms that are associated with ageing.

During the past decade, there were attempts to classify mechanisms of ageing. Examples of these efforts were the pillars of ageing (**left**) and the hallmarks of ageing (**right**).

Ageing is proven to be a highly plastic process. Genetic manipulations and pharmacological interventions could significantly extend lifespan of model organisms. For example, according to the GenAge database of ageing-related genes, more than 2,000 genes are associated with longevity (Tacutu *et al.*, 2018). Likewise, the DrugAge database collects more than 1,000 drugs and compounds that increase lifespan of model organisms (Barardo *et al.*, 2017). In addition to lifespan extension, many of these interventions could also improve healthspan (the period of life spent in good health), at least in model organisms. For instance, rapamycin, a prominent inhibitor of the mechanistic Target of Rapamycin (mTOR) pathway, could extend lifespan of several mice strains (Harrison *et al.*, 2009). Rapamycin treatment also protected mouse models from a wide range of age-related diseases (ARDs), such as cancer, cardiovascular diseases, and neurological disorders (Partridge *et al.*, 2020). Therefore, these interventions thus hold great promise to play a significant role in improving human health in the coming decades. Together, as ageing is one of the major challenges in biomedical science nowadays and our understanding of ageing processes is still in its infancy, research on ageing is urgently needed to unravel the mechanisms underlying ageing processes. Eventually, this knowledge will be crucial to developing interventions to retard or reverse ageing.

Indeed, increased age is also a leading risk factor for several ARDs, including cardiovascular diseases, neurodegenerative diseases, and cancer (Li *et al.*, 2021b). As ageing societies

worldwide continue to rise, it is imperative to prevent and develop novel therapeutic strategies to treat ARDs to support healthy ageing. Because current research and medical practice tend to focus on a single disease, ignoring the fact that several ARDs may share common underlying processes of ageing, it is necessary to understand the ageing processes and how they contribute to ARDs. Several age-related processes have been linked to numerous ARDs. For example, cellular senescence that increases with ageing contributes to atherosclerosis, diabetes, and cancer (Childs *et al.*, 2015). I discuss more the roles of cellular senescence in cancer in section **1.1.5**. In addition, chronic inflammation that occurs during ageing (inflammaging) could also contribute to various age-related pathologies through inflammatory cytokines. In fact, it has been suggested that ageing and ARDs are continuous processes with different rates/speeds, without a clear boundary between them, further highlighting the need to understand the underlying biological processes of ageing that may eventually contribute to ARDs (Franceschi *et al.*, 2018a).

1.1.2 Cancer incidence by age

Perhaps one of the most obvious connections between ageing and cancer is that fact that cancer incidence rate increases exponentially with age in most, albeit not all, cancer types (de Magalhaes, 2013). As shown in **Figure 1.2**, the data from Cancer Research UK (www.cancerresearchuk.org) reported that cancer incidence significantly rises after the middle age in most common cancers, such as breast, colorectal, lung, prostate, bladder, and brain cancers. It remains largely unknown, however, why cancer incidence declines in the very old individuals. An obvious exception is found in testicular germ cell tumours, where the incidence rate reaches its peak in young men, about the age of 25-35 years old, and sharply declines after that. The reasons why this unique pattern of age-related incidence in testicular germ cell tumours are, however, unclear. Thyroid cancer present another pattern of age-related incident rate, where incident rate increases from adolescent to middle age, and then plateau or slightly decreases after the age of 60.

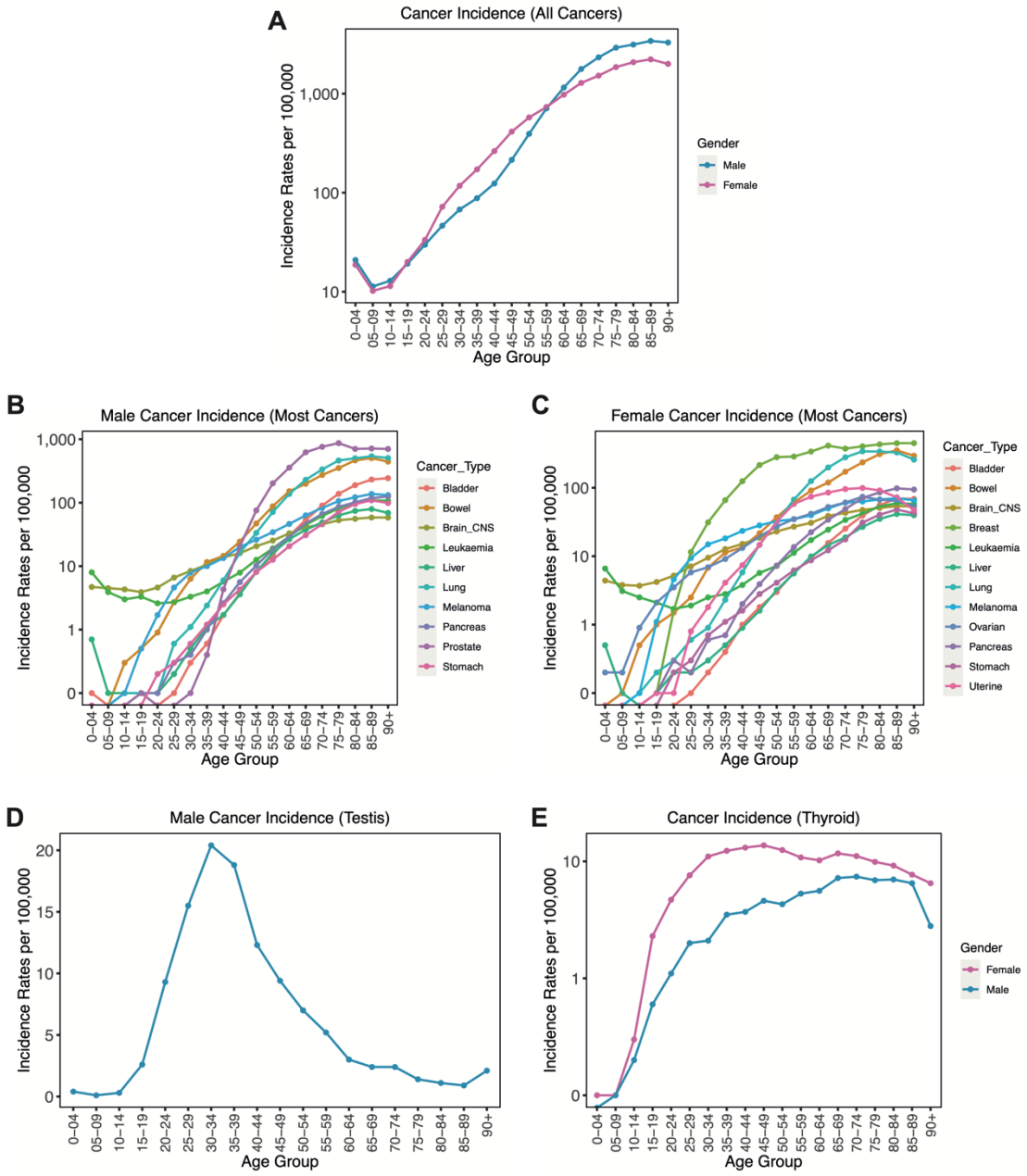


Figure 1.2 The incidence of cancer by age.

Cancer incidence rate per 100,000 male and female population from the UK (2016 – 2018) for **(A)** all cancer types, **(B-C)** most male and female cancer types in which cancer incidence increases with age, **(D)** testicular germ cell cancer, and **(E)** thyroid cancer. This figure is based on a graphic created by Cancer Research UK.

1.1.3 Ageing and cancer: The shared and opposite mechanisms

At first glance, ageing and cancer may seem like contradictory processes. Ageing is related to the decline in cellular and tissue functions, while cancer can be viewed as a gain of active cellular mechanisms such as cell proliferation. However, ageing and cancer share some common underlying hallmarks, particularly the damage to the genome and epigenome (Serrano & Blasco, 2007; Aunan *et al.*, 2017). These alterations have been thought to influence normal ageing processes that may promote carcinogenesis. Below, I will briefly recap the hallmarks of cancer and ageing. I will discuss the connection between ageing and cancer through similar and opposite hallmarks. In addition, I will also refer to other critical biological processes that are shared between ageing and cancer.

1.1.3.1 The hallmarks of cancer and the hallmarks of ageing

The hallmarks of cancer refer to the biological capabilities that cancer cells acquire during the multistep carcinogenesis and progression (Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011; Hanahan, 2022). The six original hallmarks of cancer were first proposed by Douglas Hanahan and Robert A. Weinberg in 2000 (Hanahan & Weinberg, 2000). These hallmarks include 1) self-sufficiency in growth signals, 2) insensitivity to anti-growth signals, 3) evading apoptosis, 4) limitless replicative potential, 5) sustained angiogenesis, and 6) tissue invasion and metastasis. In 2010, a revised version of the hallmarks of cancer was published, adding two emerging hallmarks: reprogramming of energy metabolisms and evading immune destruction (Hanahan & Weinberg, 2011). Furthermore, they also propose two enabling characteristics, genomic instability and tumour-promoting inflammation, that make the acquisition of these hallmarks possible. Recently, the latest version of the hallmarks of cancer has been released, adding two emerging hallmarks and two enabling characteristics (Hanahan, 2022). The two new emerging hallmarks are unlocking phenotypic plasticity and senescent cells. And two new enabling characteristics that allow cancer cells to adopt these phenotypes are nonmutational epigenetic reprogramming and polymorphic microbiomes. Overall, the hallmarks of cancer conceptualise key ideas in cancer biology, providing a useful concept for the cancer research community.

Following the success of the hallmarks of cancer to help conceptualise fundamental mechanisms underlying cancer, the hallmarks of ageing were proposed in 2013 in an attempt

to identify and classify the cellular and molecular processes associated with ageing (**Figure 1.1 right**) (Lopez-Otin *et al.*, 2013). The nine hallmarks of ageing comprise 1) genomic instability, 2) telomere attrition, 3) epigenetic alterations, 4) loss of proteostasis, 5) deregulated nutrient sensing, 6) mitochondria dysfunction, 7) cellular senescence, 8) stem cell exhaustion, and 9) altered intercellular communication. Although having recently been criticised that they do not provide an explanatory paradigm to understand the mechanistic causes of ageing (Gems & de Magalhaes, 2021), the hallmarks of ageing are helpful to provide an overview of the molecular processes that occur during ageing. Indeed, some of these hallmarks are overlapped with the hallmarks of cancer, such as genomic instability. Below, I outline the similar and opposite hallmarks between ageing and cancer and discuss shared biological pathways between these two intricated processes.

1.1.3.2 The shared and divergent processes between ageing and cancer

1.1.3.2.1 Genomic instability

One of the major shared hallmarks between ageing and cancer is genomic instability (Audia & Campbell, 2016; Zabransky *et al.*, 2022). Various forms of endogenous and exogenous stress, such as DNA replication errors, free radicals, and UV radiation, lead to DNA damage and genomic instability of the cells. As a result, somatic mutations, either at a single-nucleotide level or large-scale structural variations, occur in normal cells (Martincorena & Campbell, 2015; Martincorena, 2019). These mutations enable normal cells to acquire cancer traits, such as uncontrolled cell proliferation. Therefore, genomic instability is considered an "enabling hallmark" of cancer (Hanahan & Weinberg, 2011). Indeed, as genetic damage accumulates over time, somatic mutations and genomic instability increase with age (Li *et al.*, 2021a; Moore *et al.*, 2021). As I elaborate on later in section **1.2.1**, recent studies reported an increase in somatic mutations with age in most, if not all, non-cancerous human tissues (Li *et al.*, 2021a; Moore *et al.*, 2021). Likewise, other forms of DNA damage, such as chromosomal aneuploidies and copy-number variations, although rare, increase with age in non-cancerous individuals (Li *et al.*, 2021a). Not surprisingly, this age-related rise in genomic damage and somatic mutations has been viewed as a primary mechanism explaining an age-related increase in cancer incidence (Aunan *et al.*, 2017).

While it is generally accepted that somatic mutations play an important role in cancer progression, it is largely unclear if somatic mutations contribute to ageing phenotypes. The disruption in the DNA repair pathway can cause progeroid syndromes, such as Werner syndrome, caused by mutations in a helicase gene *WRN* (Lopez-Otin *et al.*, 2013; Vijg & Dong, 2020). Werner syndrome's patients show signs of accelerated ageing, like hair greying and loss of organ function, with a significant reduced lifespan. However, recent studies reported that individuals with inherited defects in DNA polymerases, such as those carrying *POLE/POLD1* and *MUTYH* germline mutations, do not show signs of premature ageing, although they contain a higher somatic mutation load and increase risk of cancer (Robinson *et al.*, 2021a; Robinson *et al.*, 2021b). Thus, the interplay between somatic mutations and ageing is currently far from conclusive.

1.1.3.2.2 Telomere attrition

Another form of genomic damage that may be important in the context of ageing and cancer is telomere attrition. Telomeres are repetitive sequences (5'-TTAGGGG-3') at the ends of chromosomes and are thought to protect these chromosome ends from DNA degradation (Aunan *et al.*, 2017; Rossiello *et al.*, 2022). As telomere is shortened each time cell divide, normal cells can only proliferate in a limited number of rounds before DNA damage response triggers the senescence programme and cell cycle arrest occurs (discuss more on cellular senescence in section **1.1.5**). Therefore, it is widely accepted that telomere shortening evolved as a tumour-suppressor mechanism. Cancer cells, however, express telomerase, an enzyme that can elongate the telomeres, enabling them to overcome cell cycle arrest and proliferate indefinitely (Aunan *et al.*, 2017; Rossiello *et al.*, 2022).

As expected, population studies reported an inverse relationship between age and telomere length (Serrano & Blasco, 2007; Aunan *et al.*, 2017; Demanelis *et al.*, 2020). Importantly, short telomere length has also been associated with many ARDs such as diabetes, cardiovascular diseases, and osteoarthritis (Rossiello *et al.*, 2022). It has been proposed that telomere shortening could contribute to ageing and age-related pathologies by driving cellular senescence. Senescent cells could secrete various cytokines that facilitate the age-related loss of tissue functions. Furthermore, telomere shortening can impair tissue regeneration by compromising stem cell functions. Mechanistically, telomeres shorten during cell division and activate a chronic low-level DNA damage signal, leading to a slight elevation of p53 activity.

Thus, this activation of the stress response pathway predisposes stem cells to cell cycle arrest, impairing the regenerative potential of these stem cells (Campisi & Yaswen, 2009). It has been reported that telomerase-deficient mice age prematurely due to organ defects that could be driven by impaired tissue regeneration potential of stem-cell compartments (Serrano & Blasco, 2007).

Telomerase overexpression experiments have been used to investigate if telomere extension could counteract ageing phenotypes. Indeed, telomerase overexpression in mice improves processes such as wound healing, but it also activates cancer progression (Gonzalez-Suarez *et al.*, 2001). A combined overexpression of telomerase and tumour suppressor genes, such as TP53, results in cancer-free mice with increased lifespan (Tomas-Loba *et al.*, 2008). In addition, a small molecule activator of telomerase increases healthspan in adult/old mice without increasing cancer incidence (Bernardes de Jesus *et al.*, 2011). However, the relevance of telomere attrition in humans and the clinical values remain unclear.

1.1.3.2.3 Epigenetic alterations

Epigenetics refers to processes that result in changes in gene expression without altering the DNA sequence, including DNA methylation, histone modification, chromatin remodelling, and non-coding RNAs (Allis & Jenuwein, 2016). In addition to genomic damages, epigenetic alterations have been implicated in both ageing and cancer, and multiple changes in epigenetic regulation are shared between both processes (Pal & Tyler, 2016; Sen *et al.*, 2016; Aunan *et al.*, 2017; Zabransky *et al.*, 2022). Therefore, epigenetic alterations during normal ageing may impact carcinogenesis and cancer progression (Zabransky *et al.*, 2022). For instance, DNA methylation, which is regarded as a gene suppression mechanism, at promoters of tumour suppressor genes increases with age in normal cells (So *et al.*, 2006). This observation suggests that some normal cells may be "primed" for an oncogenic transformation. As current epigenetic studies in ageing and cancer usually exploit large-scale omics technologies, I review the current knowledge of this topic in section **1.2.3**.

1.1.3.2.4 Immune system

Ageing is associated with a decline in immune system function (immunosenescence) and chronic and persistent inflammation (inflammaging) (de Magalhaes *et al.*, 2009; de Magalhaes,

2013; Fulop *et al.*, 2017; Palmer *et al.*, 2021). Immunosenescence is linked with a decrease in immune cell ability to detect and eliminate cancer cells (known as immunosurveillance) (Pawelec, 2017), while inflammaging is associated with carcinogenesis and cancer progression (Leonardi *et al.*, 2018). Therefore, age-related decline in immune system function may allow cancer to develop.

It is also widely recognised that inflammation plays a vital role in oncogenesis and is one of the hallmarks of cancer (Hanahan & Weinberg, 2011). Meta-analyses revealed the upregulation of inflammatory gene signatures during ageing across tissues and species (de Magalhaes *et al.*, 2009; Palmer *et al.*, 2021). Furthermore, inflammatory cytokines such as interleukin-6 (IL-6) and tumour necrosis factor (TNF) increase with age. Senescent cells can release these cytokines as components in senescence-associated secretory phenotypes (SASPs) (Leonardi *et al.*, 2018). A recent study reported that enhanced inflammation and a decline in tumour suppressor pathways are associated with lung tumours in aged mice (Parikh *et al.*, 2018). Together, age-related changes in the tissue microenvironment may facilitate cancer development through inflammation.

Notably, recent advances in single-cell genomic technologies facilitate the study of age-related changes in the immune cell population (Mogilenko *et al.*, 2021). For example, a recent study by Huang *et al.* reported that ageing is related to an increase in the proportion of monocytes and reduced naïve cytotoxic ($CD8^+$) and helper ($CD4^+$) T cells in the blood (Huang *et al.*, 2021). Thymic involution, the shrinkage of the thymus with age, could partly be responsible for reducing T cell production and altering T cell antigen receptor (TCR) diversity. Thus, thymic involution contributes to the rise in cancer incidence with age (Palmer *et al.*, 2018). Most immune cell types alter substantially with age, including the shifts in subpopulations of each cell type (Berben *et al.*, 2021; Mogilenko *et al.*, 2021). For cells in the innate immune system, the total number of natural killer (NK) cells increases with age, with an increase in mature and a decrease in immature cell subpopulations. Moreover, the proinflammatory monocytes ($CD16^+$) increase, while the phagocytic monocytes ($CD16^-$) decrease in older individuals. Cells in the adaptive immune system, including B and T cells, also shift in numbers and subpopulations. In addition to naïve T cells, naïve B cells are also reduced. This decline in naïve cells was accompanied by increased memory cells and regulatory T cells (T_{reg}). As cancer cells must avoid the recognition and destruction by immune cells (Hanahan & Weinberg, 2011), these shifts in the immune cell population during ageing are likely to impair the ability

to recognise antigens and decrease the anti-tumour response, creating a protumour microenvironment (Berben *et al.*, 2021; Mogilenko *et al.*, 2021).

1.1.3.2.5 Longevity and cancer pathways

Several genetic, dietary, and pharmacology interventions that affect ageing could also impact cancer. For example, rapamycin treatment extends lifespan and decreases cancer incidence and slows cancer progression in mice (Wilkinson *et al.*, 2012; Neff *et al.*, 2013). Caloric restriction that is known to increase lifespan also reduces cancer incidence (Meynet & Ricci, 2014). These observations suggest overlapping mechanisms between ageing and cancer. Indeed, several components in the longevity pathway are also cancer-related genes. There is a significant overlap (odds ratio = 10.19, *p*-value = 9.09 x 10^{-38}) between ageing-related genes from the GenAge database (Tacutu *et al.*, 2018) and cancer driver genes from Integrative OncoGenomics (IntOGen) (Martinez-Jimenez *et al.*, 2020) (**Figure 1.3**).

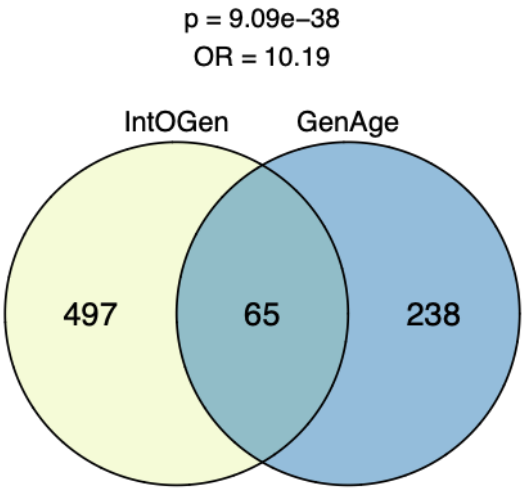


Figure 1.3. Overlap analysis between ageing-related genes and cancer driver genes. Fisher's exact test was performed to investigate the overlap between cancer driver genes from the intOGen database and ageing-related genes from the GenAge database.

One of the most well-studied longevity pathways is the insulin/IGF-1 signalling (IIS) pathway (Campisi & Yaswen, 2009). Insulin and IGF-1 signalling usually relay signals through the same downstream activations, including AKT, mTOR, RAS, and FOXO (Aunan *et al.*, 2017). Caloric restriction in mice showed a decrease in IIS activity (Fontana *et al.*, 2010). In addition, the reduction of IIS in mice results in a slight increase in lifespan (Campisi & Yaswen, 2009). In humans, population genetic studies revealed that mutations in the IGF-1 receptor gene and

FOXO3A gene were enriched in centenarians (Suh *et al.*, 2008; Willcox *et al.*, 2008). Together, these studies suggest that the IIS pathway is associated with ageing and longevity. The IIS pathway has been related to activating cell proliferation and, thus, has been implicated in cancer (Djiogue *et al.*, 2013). For example, the FOXO3a transcription factor is frequently inactivated in cancer cell lines (Dubrovska *et al.*, 2009; Liu *et al.*, 2018a). Therefore, decreasing IIS and activating FOXO could, in principle, increase longevity and decrease cancer.

Another important pathway involved in both ageing and cancer is the mTOR pathway. Widely regarded for its central roles in nutrient sensing, cell growth, and metabolism, mTOR signalling is upregulated and involved in the development of many cancers, including colorectal, prostate, and breast cancers (Zou *et al.*, 2020). Not surprisingly, several mTOR inhibitors are being used in clinical trials to treat various cancer types (Hua *et al.*, 2019). Regarding ageing, downregulation of the mTOR pathway using genetic and pharmacology approaches extends lifespan in model organisms, including yeasts, worms, fruit flies, and mice (Papadopoli *et al.*, 2019). In summary, shared important genes and signalling pathways between ageing and cancer suggest an intimate relationship between the two processes. While there is a possibility to use similar interventions to delay ageing and prevent cancer, these genes/pathways often play multiple roles in a context-dependent manner. Thus, further studies are required to identify target genes and pathways for extending lifespan and protecting from cancer.

1.1.4 Cellular senescence, ageing, and cancer

1.1.4.1 Cellular senescence: an overview

Cellular senescence refers to a stage of a cell in which irreversible cell-cycle arrest occurs. Senescent cells were first described by Hayflick and Moorhead in 1961 after the observation that human cells stop proliferating after a finite number of rounds (Hayflick & Moorhead, 1961). This phenomenon was later referred to as “replicative senescence”, which is generally caused by telomere shortening following cell divisions. Other types of cellular stress, such as oxidative stress and chemotherapeutic drugs, could also induce cellular senescence via DNA damage (Di Micco *et al.*, 2021). Furthermore, the activation of oncogenes such as *BRAF* and the RAS gene family, or the inactivation of tumour suppressor genes, such as *PTEN*, could lead to cellular senescence. This type of senescence is often referred to as “oncogene-induced senescence” (Hernandez-Segura *et al.*, 2018).

When facing different types of stress, a cell then activates the DNA damage response (DDR) pathway (Di Micco *et al.*, 2021). In brief, DNA damage activates DDR through the phosphorylation of several kinases such as ATM and ATR, which in tune start CHK2 and CHK1, respectively. Finally, these signalling components activate the p53 tumour suppressor gene. This gene induces cell-cycle arrest by activating cell cycle inhibitor $p21^{WAF1/CIP1}$ (*CDKN1A*). In addition to the p53/p21 pathway, growth arrest can also be activated by activating $p16^{INK4a}$ (*CDKN2A*). $p^{16INK4a}$ activates RB family proteins, which repress E2F family transcription factor activity, which is important in cell proliferation. These mechanisms prevent cells with genomic damage from progressing into cancer cells. Thus, the senescence response is widely regarded as a tumour suppressive mechanism.

In addition to the cell-cycle arrest, several features are also used to characterise senescent cells (Hernandez-Segura *et al.*, 2018). Morphologically, senescent cells generally flatten, enlarge, and accumulate lysosomes and mitochondria. They are often detected using senescence-associated β-galactosidase (SA-βGal). Furthermore, most senescent cells display profound changes in chromatin organisation and epigenomes. Importantly, some senescent cells can present senescence-associated heterochromatin foci (SAHFs), intense nuclear foci enriched in repressive epigenetic marks, such as tri-methylated histone 3 at lysine 9 (H3K9me3), heterochromatin protein 1 (HP1), histone variant macroH2A and histone co-chaperones HIRA and ASF1A. This repressive epigenome may silence certain cell-cycle genes.

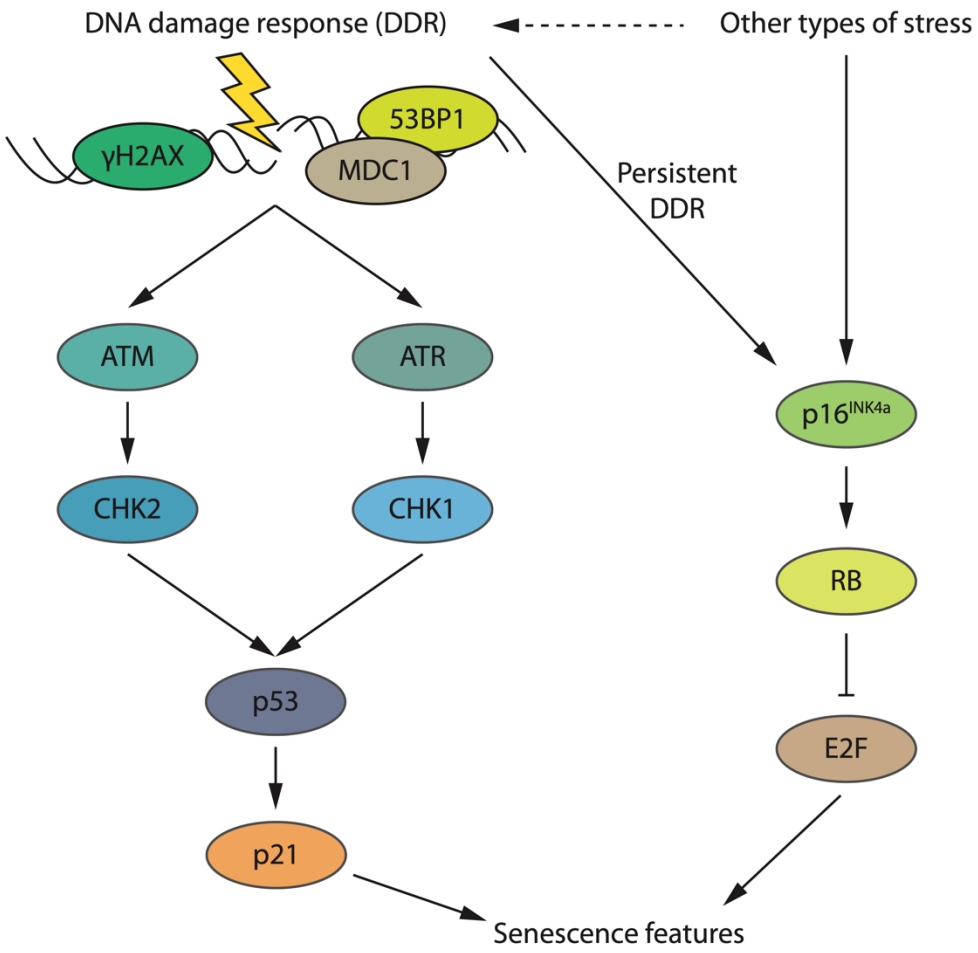


Figure 1.4. A simplified senescence pathway.

A diagram represents cellular senescence pathway, which is triggered by stress such as DNA damage.

Another important characteristic of senescent cells is an ability to secrete SASPs, including cytokines, chemokines, angiogenic factors, and extracellular matrix (ECM) proteinases (Hernandez-Segura *et al.*, 2018; Gorgoulis *et al.*, 2019). Major components of SASPs include proinflammatory cytokine IL-6, CXC chemokine ligand 8 (CXCL8 or IL-8), and monocyte chemoattractant protein 1 (MCP1). Furthermore, factors involved in ECM remodelling are also found in SASPs, such as matrix metalloproteinases (MMPs), serine/cysteine proteinase inhibitors (SERPINs), and tissue inhibitors of metalloproteinases (TIMPs) (Hernandez-Segura *et al.*, 2018; Faget *et al.*, 2019; Gorgoulis *et al.*, 2019). In addition, senescent cells could secrete non-protein molecules such as reactive oxygen species (ROS) and transported ions (Hernandez-Segura *et al.*, 2018). Senescent cells can further communicate with local and a more distal environment via extracellular vesicles, such as exosomes containing proteins and DNA (Takasugi, 2018; Borghesan *et al.*, 2019). The SASP composition and strength could also

vary substantially by cell type, duration of senescence, and the initial stimuli. As evidenced by a recent single-cell RNA-sequencing (scRNA-seq) study, the expression of SASP genes varies considerably between cells (Wiley *et al.*, 2017). Due to the complexity of the SASP, a recent effort has used proteomics to characterise soluble proteins and exosomal cargo SASP factors from multiple inducers and cell types, resulting in the "SASP atlas" (www.SASPAtlas.com) (Basisty *et al.*, 2020).

SASPs can be regulated at various levels. Notably, the DDR pathway mediates a transcriptional programme of key SASP genes, including inflammatory cytokines, through the proinflammatory transcription factor NF-κB. More recently, the cGAS-STING pathway has also been implicated in activating the transcription of SASP genes, potentially by activating NF-κB (Yang *et al.*, 2017). Other important transcription factors regulating SASP genes include GATA4 and C/EBPβ. Furthermore, in a PTEN-loss induce senescence, such as in the PTEN-deficient prostate cancer mouse model, the JAK2-STAT3 pathway activates a different set of SASP genes, mainly immunosuppressive factors (Toso *et al.*, 2014). In addition, it has been reported that mTOR signalling controls IL-1α translation to regulate the SASP (Laberge *et al.*, 2015). At the epigenetic level, repressive histone marks such as di-methylated histone 3 at lysine 9 (H3K9me2) are decreased at the promotors of core SASP genes, associated with the activation of transcription (Takahashi *et al.*, 2012). Several epigenetic modifiers such as bromodomain-containing protein 4 (BRD4), histone-lysine *N*-methyltransferase 2A (MLL1), and high-mobility group B protein (HMGB1) also facilitate SASP production (Faget *et al.*, 2019; Di Micco *et al.*, 2021).

1.1.4.2 Cellular senescence and ageing

Although the lack of reliable markers posits a challenge in detecting senescent cells in tissues and organs, the use of multiple potential senescent markers provides evidence of the accumulation of senescent cells in primates (Herbig *et al.*, 2006; van Deursen, 2014) and potentially humans (Tuttle *et al.*, 2020). These markers include a high level of $p16^{Ink4a}$, p21, IL-6, and SA-βGal (van Deursen, 2014). Different organs may display different levels of these markers (Idda *et al.*, 2020; Tuttle *et al.*, 2021). Senescent cells and SASPs have been reported to involve in various biological processes, including development (Munoz-Espin *et al.*, 2013), wound healing (Demaria *et al.*, 2014), and cancer progression (Campisi, 2013). Furthermore, SASPs could affect the local tissue microenvironment, mediating various age-related

pathologies. Senescent cells are thought to promote age-related tissue and organ deterioration through multiple mechanisms (van Deursen, 2014). First, senescent stem cells impair tissue homeostasis and regeneration (Brack *et al.*, 2007). In addition, SASPs secreted from senescent cells can disrupt the stem cell niche in a non-cell-autonomous fashion (Gnani *et al.*, 2019). Next, protease components in SASPs could disrupt tissue architecture, such as degenerate ECM. Other SASP components, such as proinflammatory cytokines IL-6, may contribute to a chronic tissue inflammation (inflammaging) related to various ARDs such as cancer and cardiovascular (Ferrucci & Fabbri, 2018; Franceschi *et al.*, 2018b). Finally, paracrine signalling from senescent cells may spread senescent phenotypes to their neighbouring cells (Nelson *et al.*, 2012; Gnani *et al.*, 2019).

Because of the potential roles of senescent cells contributing to ageing pathologies and ARDs, targeting cellular senescence has become an attractive strategy to promote healthy ageing (Gasek *et al.*, 2021). An early proof-of-concept study reported that clearance of senescent cells in the progeroid $BubR1^{H/H}$ mouse model by removing cells expressing high levels of p16 could alleviate age-related phenotypes (Baker *et al.*, 2011). Later, numerous mouse models of accelerated senescence and ageing have been developed and tested for assessing the role of cellular senescence in ageing (Gorgoulis *et al.*, 2019; Di Micco *et al.*, 2021; Gasek *et al.*, 2021). For instance, clearance of senescent cells attenuates the development of post-traumatic osteoarthritis in the p16-3MR transgenic mouse model (Jeon *et al.*, 2017). In addition, the removal of senescent cells was shown to mitigate the premature ageing of the haematopoietic stem cells in mice (Chang *et al.*, 2016). In accordance with these previous studies, Xu *et al.* reported that the administration of senolytic drugs, drugs that eliminate senescent cells, improved physical function and increased lifespan in mice (Xu *et al.*, 2018).

Together, targeting cellular senescence could be a promising strategy for combating ageing and age-related diseases. Since the field is still in its early days, it will be beneficial from the currently ongoing clinical trial studies (Gasek *et al.*, 2021).

1.1.4.3 Dual roles of cellular senescence in cancer

Because of its phenotype in growth arrest, cellular senescence is widely regarded as a potent tumour suppressive mechanism. As discussed above, various stress and DDR generally activate senescence growth arrest through multiple tumour suppressor genes, such as p53/p21 and $p16^{INK4a}$. As such, the growth arrest suppresses the development of tumours by preventing the cells carrying damaged DNA to proliferate and further acquire oncogenic mutations (Campisi, 2013). In contrast, increasing evidence shows that senescent cells can fuel tumourigenesis. In a transgenic mouse model, selective elimination of senescent cells decreases spontaneous tumour formation (Baker *et al.*, 2016). Senescent cells support tumour initiation and progression by secretion of pro-tumorigenic and angiogenic factors as SASP components, such as MMP3 and vascular endothelial growth factor (VEGF) (Krtolica *et al.*, 2001; Coppe *et al.*, 2006; Liu & Hornsby, 2007). IL-6, a common SASP factor, can promote the proliferation of various cancer cell types in mouse models, including skin (Lederle *et al.*, 2011), breast (Hartman *et al.*, 2013), prostate (Rojas *et al.*, 2011), and lung (Song *et al.*, 2011). Furthermore, it has been reported that senescent cell-derived exosomes could promote cancer cell proliferation (Takasugi *et al.*, 2017).

In addition to cell proliferation, SASP factors can induce various other oncogenic processes. Angiogenesis, the growth of blood vessels to supply nutrients to tumours, can be induced by VEGF, a component of SASPs (Coppe *et al.*, 2006; Oubaha *et al.*, 2016; Gonzalez-Meljem *et al.*, 2018). SASPs can promote tumour cell invasion by inducing epithelial-to-mesenchymal transition (EMT) (Laberge *et al.*, 2012). This characteristic is also a common feature of cancer metastasis (Hanahan & Weinberg, 2011). Indeed, SASP factors such as MMP facilitate ECM remodelling, which can, in turn, influence metastasis (Coppe *et al.*, 2010; Malaquin *et al.*, 2013; Faget *et al.*, 2019). Interestingly, several studies reported that SASP could promote resistance to anticancer therapy, either by protecting cancer cells from chemotherapy (Kaur *et al.*, 2016) or by promoting stem cell-like EMT phenotypes in cancer (Milanovic *et al.*, 2018).

While cellular senescence prevents cells containing genomic damage from transforming into cancer cells, SASPs secreted from senescent cells can promote cancer progression in various mechanisms. Thus, targeting cellular senescence could be a promising strategy for cancer treatment, especially in elderly patients.

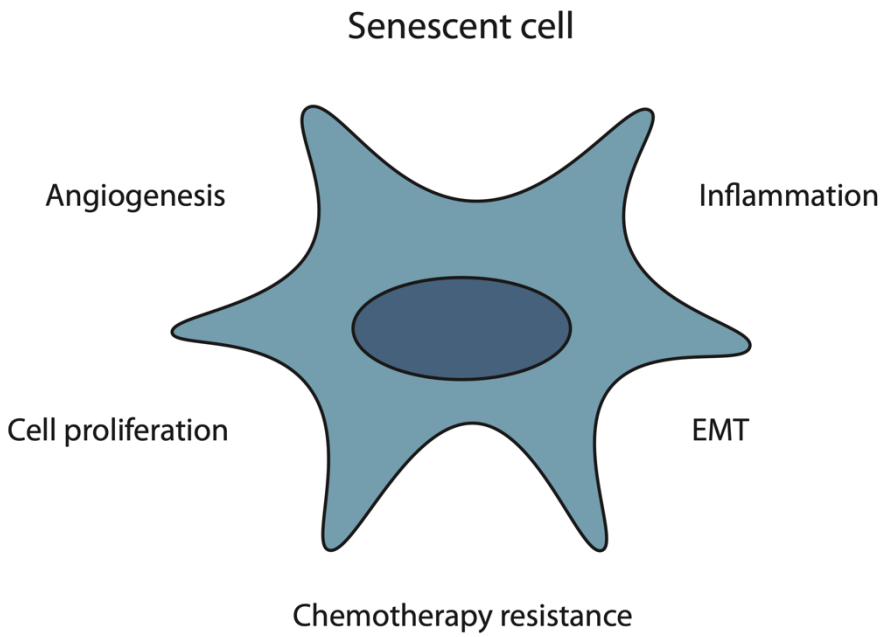


Figure 1.5. Roles of cellular senescence in promoting cancer progression.
SASPs secreted from senescent cells can influence cancer progression through various mechanisms, such as increasing inflammation, promoting angiogenesis, facilitating EMT, inducing cell proliferation, and promoting chemotherapy resistance.

1.2 Ageing and cancer in the postgenomic era

During the past two decades, large-scale "omics" analyses have played an important role in almost, if not all, areas of biological sciences. In particular, the advent of next-generation sequencing has transformed research in an unprecedented way (Goodwin *et al.*, 2016). For example, targeted sequencing, whole-exome sequencing (WES), and whole-genome sequencing (WGS) have been particularly important in cancer research, pinpointing mutational events in the cancer genome (Stratton *et al.*, 2009; Wong *et al.*, 2011; Garraway & Lander, 2013). Gene expression microarray and RNA sequencing (RNA-seq) enable a high-throughput transcriptome-wide analysis to identify differentially expressed genes between biological conditions, such as normal control and diseases (Stark *et al.*, 2019). Furthermore, gene expression signatures derived from transcriptomic assays can be used as biomarkers. Likewise, various methods have been developed to study the gene regulatory logic through epigenetic mechanisms, to better understand mechanisms controlling gene expression (Zentner & Henikoff, 2014).

Following these technological advances, several large-scale international consortium efforts have been established with the goal to systematically characterise molecular alterations across cancer types. The most notable projects include The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research *et al.*, 2013b), International Cancer Genome Consortium (ICGC) (International Cancer Genome *et al.*, 2010), and, more recently, the Pan-Cancer Analysis of Whole Genome (PCAWG) (Consortium, 2020). With these efforts, our understanding of cancer driver genes has substantially increased, with currently more than 568 cancer driver genes have been identified (Source: IntOGen) (Martinez-Jimenez *et al.*, 2020). In addition, other types of molecular data, such as gene expression and DNA methylation, and clinical data are also available, facilitating multi-layers integrative data analyses for cancer.

Interestingly, accumulating evidence has indicated the clonally expanded clones in several non-cancerous tissues that contain mutations in cancer driver genes (Yizhak *et al.*, 2019; Kakiuchi & Ogawa, 2021; Li *et al.*, 2021a; Moore *et al.*, 2021). These cancer driver-harbouring clones typically show an age-related increase pattern, suggesting the potential link between ageing, somatic mutations, and carcinogenesis. Furthermore, changes in transcriptomic and epigenetic landscapes of tissues during ageing have been extensively studied, thus likely to provide clues on how ageing-related alterations in the tissue microenvironment might enable cancer progression (Fane & Weeraratna, 2020). In addition, as discussed above, several age-related phenomena, such as cellular senescence and the ageing immune system, may also contribute to an increase in cancer incidence in the elderly.

In this section, I discuss examples of the use of high throughput "omics" technologies to reveal the mechanisms underlying ageing processes that may influence cancer initiation and progression. Firstly, I focus on the growing evidence of somatic mutations in non-cancerous human tissues, which is particularly valuable to increase our understanding of cancer initiation. I further discuss age-related changes in tissues through transcriptomic and epigenetic lenses. The changes that might facilitate cancer progression.

1.2.1 Age-related somatic mutations across tissues

As mentioned above, one of the common hallmarks between ageing and cancer is genomic instability, including the accumulation of somatic mutations. Studies on the cancer genome revealed an increase in somatic point mutation burden with age. Spontaneous deamination of 5-methylcytosine to thymine at CpG dinucleotides is an endogenous process driving the age-associated mutation load and is often referred to as a clock-like mutational process. It is well accepted that tumours arise from the expansion of positively selected clones fuelled by oncogenic events such as mutations in oncogenes and tumour suppressor genes. However, detection of clonal evolution in normal tissues has not been possible previously due to the technical limitation to detect much smaller clones than those in cancer. Recently, advances in DNA sequencing technologies allow a detailed investigation of the somatic mutation landscape in non-cancerous tissues. In this section, I revisit an emerging body of evidence implicating an accumulation of age-related somatic mutant clones in blood and normal solid tissues and their implications for cancer development.

1.2.1.1 Age-related clonal haematopoiesis and blood cancer

Various types of somatic mutations – single nucleotide variants (SNVs), small insertions and deletions (indels), and large-scale structural variations (SVs) – accumulate in tissues throughout life. Intrinsic processes, such as errors during DNA replication and repair, and extrinsic processes like exposure to environmental factors, contribute to this somatic mutation accumulation (Vijg & Dong, 2020). Although most mutations do not affect the fitness of a cell, some mutations could have a positive or negative impact on a cell. When a cell acquires a mutation that confers its selective advantage, it can expand into a clone through Darwinian evolution (Martincorena, 2019). Probably due to its accessibility, early studies of the clonal expansion process in non-cancerous tissues come from studies in the blood system. Clonal haematopoiesis (CH) is a general term that refers to the expansion of clones from a single haematopoietic stem cell, which is often driven by mutations in known blood cancer driver genes (Jaiswal & Ebert, 2019). Another related term, clonal haematopoiesis of indeterminate potential (CHIP), refers explicitly to CH that contains mutations in haematopoietic malignancy driver genes in the blood of healthy individuals without evidence of malignancy (Silver & Jaiswal, 2019). CHIP is rarely detected in people under the age of 40. However, the frequency of CHIP increases significantly with age. CHIP is, therefore, clearly an age-related

phenomenon and is thought to be associated with haematopoietic malignancy (Jaiswal & Ebert, 2019).

In 2014, three studies employed existing large-scale WES data from blood to investigate the presence of detectable SNVs and small indels in individuals who are not known to have haematologic cancer (Genovese *et al.*, 2014; Jaiswal *et al.*, 2014; Xie *et al.*, 2014). In general, these studies' findings were largely similar. Due to differences in sequencing platform and depth and analysis pipeline, the percentage of individuals presenting CHIP is varied. However, all studies reported that the percentage of persons with CHIP rises considerably with age. For instance, Genovese et al. reported that CHIP was presented in only 1% of individuals younger than 50 years but increased significantly to 10% of those older than 65 years (Genovese *et al.*, 2014). The majority of mutations were found in three epigenetic regulator genes that are frequently mutated in haematologic cancer: *DNMT3A*, *TET2*, and *ASXL1*. Indeed, the most common base substitution was cytosine to thymine (C>T), which is the feature of age-related mutational signatures (Alexandrov *et al.*, 2015). Individuals with detectable somatic mutations are associated with reduced overall survival and have a higher chance to develop haematologic cancer, supporting the hypothesis that CHIP is a premalignant condition. Furthermore, individuals who later developed haematologic cancer had a higher mean variant allele frequency (VAF) when obtaining blood samples (Jaiswal *et al.*, 2014), indicating that the presence of a large clone is associated with a higher risk of malignancy. A subsequent study used ultra-deep sequencing to investigate CHIP with higher sensitivity and confirmed previous findings that CHIP is increasingly more common in older individuals (McKerrell *et al.*, 2015). Somatic mutations in spliceosome genes *SF3B1* and *SRSF2* were exclusively presented in persons older than 70 years. This finding suggested that clones harbouring mutations in these genes were positively selected under the microenvironment of haematopoietic ageing, highlighting the influence of ageing in shaping mutation-selection interplays to promote clonal expansion (Rossi *et al.*, 2008; McKerrell *et al.*, 2015). Interestingly, *NPM1* mutations, common among myeloid leukaemia, were undetected in the study, suggesting that *NPM1* mutations might be "gatekeepers" of leukemogenesis (McKerrell *et al.*, 2015).

In addition to SNVs and small indels, large-scale chromosomal alterations, including duplications, deletions, and copy-neutral loss of heterozygosity (LOH), have also increased with age. By analysing large-scale genome-wide single-nucleotide polymorphism (SNP) array data, two studies concurrently reported chromosomal mosaicism in individuals without

malignancy (Jacobs *et al.*, 2012; Laurie *et al.*, 2012). The frequency of samples with chromosomal alterations is less than 0.5% in persons younger than 50 but rises to about 2-3% in the elderly (over 70 years). A large proportion of regions frequently altered are also regions that show copy-number changes in haematologic cancer, including losses of chromosomes 20q, 13q, 11q and 17p, and gains of chromosomes 12 and 8 (Laurie *et al.*, 2012). Indeed, this detectable clonal mosaicism is associated with increased haematologic cancer risk (Laurie *et al.*, 2012; Jaiswal *et al.*, 2014).

CHIP is associated with an increased risk of developing haematologic cancer, which has led to an advance in early cancer detection. For example, one study showed that somatic mutations could be detected as early as ten years before acute myeloid developed (Desai *et al.*, 2018). Therefore, genome sequencing could help detect age-related clones harbouring mutations and serve as a platform for early cancer detection. Despite these advances, several important questions remain to be elucidated. For instance, although the studies mentioned above have provided concrete evidence that CHIP driven by somatic mutations in cancer driver genes have expanded throughout life, CHIP without known driver mutations has also been observed (Zink *et al.*, 2017). Furthermore, the mechanistic understanding of how CHIP emerges is also lacking. A recent study discovered that germline variants, such as a locus at *TET2* gene, interfere with haematopoietic stem cell function and increase the chance to develop CHIP (Bick *et al.*, 2020). In addition, several studies have used sequencing data to measure a fitness landscape of mutations in CHIP to better understand how these mutations contribute to CHIP and eventually to blood cancer (Watson *et al.*, 2020; Robertson *et al.*, 2021). This understanding should provide a predictive framework to stratify cancer risk. Finally, though beyond the scope of this thesis, it should be noted that CHIP has also been linked to other deficiencies, such as coronary heart disease and atherosclerosis and is associated with higher all-cause mortality (Jaiswal *et al.*, 2014; Jaiswal *et al.*, 2017).

1.2.1.2 Age-related somatic mutations in normal solid tissues

In addition to the haematopoietic system, an age-related clonal expansion has been reported in various solid tissues. Using ultradeep targeted sequencing of cancer driver genes, Martincorena and colleagues' pioneering study reported several cutaneous squamous cell carcinoma driver genes (e.g., *NOTCH1*, *NOTCH2*, *NOTCH3*, *FAT1*, and *TP53*) are positively selected in non-cancerous aged sun-exposed skin (Martincorena *et al.*, 2015). These mutations show a

signature of ultraviolet light, driven mainly by C>T substitutions. Indeed, an accumulation of somatic mutations increases with age (Hernando *et al.*, 2021). Following studies further showed that this clonal expansion of cancer driver gene-carrying clones is universal across most, if not all, solid tissues in the body, including the oesophagus (Martincorena *et al.*, 2018; Yokoyama *et al.*, 2019), colon (Lee-Six *et al.*, 2019), liver (Brunner *et al.*, 2019), lung (Yoshida *et al.*, 2020), prostate (Grossmann *et al.*, 2021), skeletal muscle (Franco *et al.*, 2018), endometrium (Moore *et al.*, 2020), urothelium (Li *et al.*, 2020b), bladder (Lawson *et al.*, 2020), and adult stem cells from liver, small intestine, and colon (Blokzijl *et al.*, 2016). These findings concordantly reported that somatic mutations increase with age, driven by endogenous clock-like mutational signatures. It should be noted that in addition to age, environmental exposure to tobacco smoking, UV light, and alcohol consumption also accelerates clonal expansion (Brunner *et al.*, 2019; Yokoyama *et al.*, 2019; Yoshida *et al.*, 2020; Hernando *et al.*, 2021). The pattern of somatic mutations in normal tissues was also confirmed by recent pan-tissue mutational studies (Li *et al.*, 2021a; Moore *et al.*, 2021). Furthermore, to obtain an overview of the somatic mutation landscape across a large population, two studies have developed a method to detect somatic mutations from the Genotype Tissue Expression (GTEx) project's RNA sequencing data (Garcia-Nieto *et al.*, 2019; Yizhak *et al.*, 2019), consisting of ~7,000 samples across ~30 tissues. It is noteworthy that this approach has several limitations, such as it can detect only mutations presented in expressed genes and those that occurred with a high VAF. Nevertheless, these studies also reported the increased clonal expansion with age across tissues. Recent scRNA-seq analyses in the human pancreas (Enge *et al.*, 2017), human B lymphocytes (Zhang *et al.*, 2019), and 23 tissues and organs across the mouse body (Tabula Muris, 2020) further confirmed the widespread accumulation of somatic mutations and clonal expansion with age across tissues.

Mutations in cancer driver genes were recurrently found in non-cancerous tissues, for example, *NOTCH1*, *NOTCH2*, *NOTCH3*, and *TP53* in the oesophagus and skin, and *AXIN2*, *STAG2*, *PIK3CA, ERBB2, ERBB3* in the colon. Unexpectedly, the frequency of somatic mutations in some of these driver genes in normal tissues may be higher than in cancer. For instance, *NOTCH1* was more frequently mutated in normal oesophagus than in oesophageal cancer (Martincorena *et al.*, 2018; Yokoyama *et al.*, 2019). This unexpected finding has suggested that *NOTCH1* mutations, even though they can promote clonal evolution, might not be necessary for oesophageal cancer or might even play a role as a protective mechanism preventing cells from oncogenic transformation (Higa & DeGregori, 2019; Martincorena,

2019; Wijewardhane *et al.*, 2021). A recent study supported this hypothesis by revealing that *Notch1* mutation promotes clonal expansion in the mouse oesophagus. However, *Notch1* loss impaired tumour growth (Abby *et al.*, 2021). Interestingly, positively expanded clones in normal oesophagus could also compete and eliminate newly formed tumour clones (Colom *et al.*, 2021). In this case, clonal competition could be viewed as a mechanism to limit carcinogenesis. Indeed, clonal competition is dynamic over time. The fitness of each clone depends on mutations it harbours but may also be modulated by the environment as well as neighbouring cells (Fernandez-Antoran *et al.*, 2019; Colom *et al.*, 2020). Thus, it is tempting to speculate that the ageing environment might, to some degree, play a role in shaping clonal selection and carcinogenesis (Laconi *et al.*, 2020).

In addition to point mutations, structural variations and copy-number alterations are frequently observed and regarded as crucial events in cancer initiation and progression (Li *et al.*, 2020c). These large-scale variants were also found in normal tissues, including the oesophagus (Yokoyama *et al.*, 2019; Li *et al.*, 2021a), liver (Brunner *et al.*, 2019), bladder (Lawson *et al.*, 2020), colon (Lee-Six *et al.*, 2019), endometrium (Moore *et al.*, 2020), and cardia (Li *et al.*, 2021a), albeit with a far less frequency than in cancer. These alterations also increased with age. Because of their very low frequency in non-cancerous tissues, it is possible that the important event to transform normal cells into malignancy could be large-scale variations. However, the studies of these structural variation landscapes in normal tissues are only starting to emerge and awaiting more research to come. In addition, it should be noted that difficulties in somatic copy-number alterations (SCNAs) and SVs detection may mean the current works are underestimating the actual number of alterations. Advances in genome sequencing technology and computational methods to reliably detect SCNAs and SVs should shed light on the large-scale somatic alterations in non-cancerous tissues. Taken together, studies using genomic and transcriptomic sequencing technologies have revealed the landscape of somatic mutations in numerous tissues and highlighted age as a major factor in mutation accumulation. The emerging evidence that somatic mutations in cancer driver genes accumulate in normal tissues without a sign of malignancy raises the question of the definition of cancer driver genes. In addition, these findings further suggest that mutations in some driver genes allow for clonal expansion. Still, malignancy transformation may require additional genetic alterations or a combination of cancer driver mutations. The combination of mutations necessary or sufficient to drive carcinogenesis is not well understood. Furthermore, the competition between clonally expanded clones and the interaction between these clones and the tissue microenvironment,

which ageing processes can modulate, remains unclear. Further studies are urgently needed to shed light on the interplay between age-related genetic mutations, clonal expansion, and clonal selection under dynamic tissue environments during ageing.

1.2.2 Age-related changes in gene expression and cancer

Ageing is accompanied by a progressive decline in the physiological function of tissues and an increased risk of age-related diseases, including cancer (Lopez-Otin *et al.*, 2013). It is, therefore, reasonable to speculate that tissue environmental changes during ageing might be beneficial for cancer progression (de Magalhaes, 2013; Fane & Weeraratna, 2020; Laconi *et al.*, 2020). Several studies have used emerging large-scale omics technologies to capture changes in different tissues across the body with age. Indeed, microarray and RNA-seq have been extensively employed to study gene expression changes associated with ageing in model organisms and humans (Valdes *et al.*, 2013). Meta-analyses of gene expression in ageing identified transcriptomic signatures of ageing, including upregulation of immune and stress response genes and downregulation of developmental and metabolic genes (de Magalhaes *et al.*, 2009; Palmer *et al.*, 2021). However, age-associated differentially expressed genes are mainly tissue-specific, with little overlap between tissues, suggesting a tissue specificity of ageing processes (Yang *et al.*, 2015; Palmer *et al.*, 2021).

Aramillo Irizar *et al.* use several ageing and age-related transcriptomic datasets across four vertebrate species, including humans, to show that ageing gene expression changes are similar to degenerative diseases, whereas opposite to cancer gene expression signatures (Aramillo Irizar *et al.*, 2018). Indeed, cancer is characterised by an uncontrolled cell division. On the other hand, ageing is associated with a decrease in stem cell division and function (de Magalhaes *et al.*, 2009; Oh *et al.*, 2014; Aramillo Irizar *et al.*, 2018). From an evolutionary point of view, cells acquired an oncogenic phenotype via somatic mutations might be able to grow and expand over surrounding cells with limited cell division ability, thus providing a selective advantage for cells harbouring oncogenic mutations, a concept called “Adaptive oncogenesis” (Marusyk & DeGregori, 2008; Laconi *et al.*, 2020). This idea is consistent with results from the mathematical modelling (Rozhok *et al.*, 2014; Rozhok & DeGregori, 2019) and the experimental studies in mice haematopoietic system, in which a decline in B-lymphocyte progenitor cell function could increase the selection for Bcr-Abl carrying cells

(Henry *et al.*, 2010; Henry *et al.*, 2015). It would be interesting to see further evidence in solid organs.

It is important to note that most studies to date used bulk RNA-seq to study gene expression changes associated with age, obscuring the cellular resolution of age-related changes in the transcriptome. The ever-increasing data generated from scRNA-seq holds a great promise to shed light on the age-related changes in tissue microenvironment and their impact on cancer. For instance, a recent study revealed age-dependent alterations in cell proportions and gene expression, loss of tumour suppressor, and change in tissue microenvironment in mouse mammary gland (Li *et al.*, 2020a). The study found the expansion with age of AV luminal progenitor cells, the potential cells of origin for triple-negative breast cancer. Furthermore, genes associated with the basement membrane were downregulated in myoepithelial cells, the change that may associate with cancer progression. Further studies employing scRNA-seq will continue to provide novel insights into how ageing-related alterations in different tissues could facilitate cancer progression.

Taken together, transcriptomic technologies have shed light on the age-related alterations in tissues, which could facilitate carcinogenesis. However, it remains to be investigated if gene expression changes in ageing and cancer are similar or different in distinct organs. Therefore, I have studied this particular question of interest and have explained the results in **Chapter 2**.

1.2.3 Age-related epigenetic alterations and cancer

Epigenetic mechanisms regulate gene expression and associate with both normal ageing and cancer. Alterations in DNA methylation, histone modifications, and chromatin structure are the hallmarks of both ageing and cancer. Notably, several epigenetic alterations are common between ageing and cancer, yet the picture is complex. Below, I review the current knowledge of epigenetic changes in ageing and how these changes may facilitate cancer initiation and progression.

Both ageing and cancer are accompanied by alterations in various epigenetic processes, including DNA methylation (Johnson *et al.*, 2012; Feinberg *et al.*, 2016; Pal & Tyler, 2016; Nebbioso *et al.*, 2018). While gene expression changes in ageing and cancer are primarily opposite, DNA methylation changes in ageing and cancer exhibit a more complex scenario.

Ageing and cancer share similar features of global hypomethylation and hypermethylation at the specific CpG islands of the promoters (Zabransky *et al.*, 2022). In ageing, the loss of global DNA methylation results in the activation of normally silenced genes and repetitive sequences in the genome (Liu *et al.,* 2011; Johnson *et al.,* 2012; Zane *et al.,* 2014). The global DNA hypomethylation in ageing is accompanied by DNA hypermethylation at specific gene promoters, such as developmental genes (Rakyan *et al.*, 2010). Intriguingly, the expression level of *DNMT1* decreases with age, which may partly explain the decline in DNA methylation during ageing (Liu *et al.*, 2011). However, the notion of age-related global hypomethylation has been challenged by recent studies. For example, using whole-genome bisulfite sequencing, the researchers found that global DNA methylation level in mouse liver is similar between young and aged mice (Cole *et al.,* 2017). Indeed, recurrent mutations in epigenetic regulators, including DNA methyltransferase enzymes (such as *DNMT1*, *DNMT3A* and *DNMT3B*) and methylcytosine dioxygenase enzymes (such as *TET1*, *TET2*, and *TET3*) have been widely observed across cancers. Two studies showed that ageing stem cells are associated with hypermethylation at promoters of the polycomb group protein target genes, such as developmental genes (Rakyan *et al.*, 2010; Teschendorff *et al.*, 2010). These genes are also frequently methylated in cancer. Due to the similarity between DNA methylation changes in ageing and cancer, it has been hypothesised that DNA methylation changes in ageing could predispose cells to malignancy transformation.

Recently, tissue-specific analyses revealed a much more complex relationship between ageing and cancer DNA methylation landscapes (Dmitrijeva *et al.*, 2018; Perez *et al.*, 2018). Using TCGA data, one study showed that hypermethylated regions in both ageing and cancer shared a similar bivalent chromatin signature. On the other hand, hypomethylated regions in ageing highly overlapped with regions marked by the active histone mark H3K4me1, while hypomethylated sequences in cancer are associated with the repressive H3K9me3 mark (Perez *et al.*, 2018). Another study reported that changes in DNA methylation in ageing do not strongly overlap with those that occur in cancer (Dmitrijeva *et al.*, 2018; Perez *et al.*, 2018). Thus, the relationship between DNA methylation changes in ageing and cancer is largely unclear.

Apart from DNA methylation, several epigenetic mechanisms have been studied in ageing and cancer. For instance, histone modification changes substantially in ageing (McCauley & Dang, 2014; Yi & Kim, 2020) and cancer (Audia & Campbell, 2016; Zhao & Shilatifard, 2019; Zhao

et al., 2021). However, studies linking histone modifications in ageing and cancer are largely lacking. One recent work studied active histone marks, including acetylation of histone 3 at lysine 27 (H3K27ac), monomethylation of histone 3 at lysine 4 (H3K4me1), and trimethylation of histone 3 at lysine 4 (H3K4me3) in aged haematopoietic stem cells (HSCs). They reported that changes in these histone marks with age in HSCs affected developmental and cancer pathways and may predispose HSCs to myeloid malignancy transformation (Adelman *et al.*, 2019). Taken together, epigenetic alterations that occur during ageing may predispose some populations of cells to carcinogenesis. Further studies are needed to shed light on this hypothesis and the molecular mechanisms linking epigenetic changes in ageing and cancer.

1.3 Cancer genomics

Cancer is a leading cause of death worldwide. It is responsible for approximately one in six deaths, with nearly 10 million deaths in 2020 (https://www.who.int/news-room/fact-sheets/detail/cancer). The most common types of cancer are breast, lung, colorectal, prostate, skin, and stomach cancers. Cancer is often referred to as a disease of the genome (Macconaill & Garraway, 2010). Various forms of exogenous and endogenous factors contribute to the genetic damage of cells, transforming normal cells into cancer cells. These factors include but are not limited to genetic background, age, sex, and carcinogens such as ultraviolet light, tobacco smoke, alcohol, virus, and bacteria. Current advances in genome sequencing have identified genetic mutations associated with diverse cancer types, transforming our understanding of the cancer genome, which is particularly important in clinical applications. This section outlines general knowledge in cancer genomics, such as types of genomic damage. I further review the current literature on the age-associated differences in the cancer genome.

1.3.1 Cancer as an evolutionary process

Cancer can be viewed as a Darwinian evolutionary process occurring among cell populations, as first formally proposed by Peter Nowell (Nowell, 1976). Analogous to Darwinian natural selection, Nowell argued that cancer arises from a stepwise process consisting of somatic mutation and clonal selection. Genetic (and epigenetic) mutations accumulate over time create phenotypic variabilities among the population of cells, whereby natural selection acts on these variations (Stratton *et al.*, 2009; Greaves & Maley, 2012; Podlaha *et al.*, 2012; Lipinski *et al.*,

2016; McGranahan & Swanton, 2017). Cells that are acquired mutations that confer an advantage to survival will overcome their neighbouring cells and thrive. Thus, cancer evolution results from the interaction between a stochastic mutation accumulation and a deterministic force of selection. Indeed, the tumour microenvironment plays a vital role in clonal selection. For example, environmental factors like immune cells, nutrient supplies, hypoxia, cancer therapeutic drugs, etc., can shape tumour evolution and allow the "fittest" clones to survive (Podlaha *et al.*, 2012; Lipinski *et al.*, 2016; McGranahan & Swanton, 2017). In addition, the process of mutation acquisition and selection are iterative, creating an intratumour heterogeneity (multiple subclones that harbour different mutations and display distinct phenotypes in a tumour), which contributes to therapeutic failure and drug resistance.

1.3.2 Genomic alterations in the cancer genome

As a material for clonal selection, genomic damages could "drive" carcinogenesis. International cancer genomic efforts used genomic technologies to identify these damages, creating a catalogue of somatic mutations in the cancer genome (Tate *et al.*, 2019; Martinez-Jimenez *et al.*, 2020). A cancer genome may harbour distinct classes of DNA sequence change, including SNVs, small indels, SCNAs, large-scale SVs, changes in chromosome number (aneuploidy), and whole-genome duplications (WGD). In this section, I provide an overview of these alterations.

1.3.2.1 Driver vs passenger mutations

As solid tumours can contain tens to thousands of genomic alterations in their genome, one of the most important challenges in the field is to distinguish aberrations that confer a selective advantage to tumours and therefore enable carcinogenesis ("driver") from those that do not contribute to cancer progression but just arise by chance ("passenger") (Stratton *et al.*, 2009; Collisson *et al.*, 2012; Vogelstein *et al.*, 2013). The primary strategy to identify driver mutations is to search for the recurrently altered genes across a large sample size, indicating that they are under positive selection and may not merely occur by chance. One method used the ratio of non-synonymous to synonymous mutations for a framework to quantify selection in the cancer genome and reported that tumours carry about four coding mutations under positive selection on average (Martincorena *et al.*, 2017). As mentioned earlier, global cancer genome projects have accelerated the discovery of driver mutations in most, if not all, cancer

types. Up-to-date lists of cancer driver mutations can be found at databases such as Catalogue of Somatic Mutations in Cancer (COSMIC) (Tate *et al.*, 2019) and IntOGen (Martinez-Jimenez *et al.*, 2020). Some known cancer driver genes are frequently mutated in diverse cancer types, such as *TP53*, *CDKN2A*, *KRAS*, and *PTEN*. However, other driver genes display a cancer-type-specific pattern (Consortium, 2020).

Driver alterations can broadly be classified as oncogenes and tumour suppressor genes (Collisson *et al.*, 2012; Vogelstein *et al.*, 2013). Oncogenes refer to genes in which alterations, such as copy number gains and missense mutations, induce upregulation and enhance cancer phenotypes. On the other hand, tumour suppressor genes generally act as negative regulators of important cancer phenotypes, such as cell proliferation. In the case of many tumour suppressor genes, pronounce oncogenic phenotypes result from driver alterations affecting both alleles (a two-hit inactivation event). For example, a recent pan-cancer study found that 77% of tumours with driver mutations in *TP53* had both alleles mutated (Consortium, 2020).

1.3.2.2 Single-nucleotide variants, mutational signatures, and short insertions/deletions

The most common type of aberrations found in the cancer genome is single-nucleotide variants (SNVs), the substitution of a single nucleotide for another. There are six classes of the SNVs: C>A, C>G, C>T, T>A, T>C, T>G. Exogenous and endogenous mutational processes cause somatic mutations. Different sources of mutations generally leave distinct mutational patterns. For instance, the spontaneous deamination of 5-methylcytosine results in C>T transition. In the early 2010s, scientists developed a method to automatically extract mutational signatures based on information on the context of each mutation and the bases immediately at 5' and 3' to each mutant base (96 possible mutated trinucleotides). This analysis yields 21 distinct original mutational signatures. Some of these mutational signatures are commonly observed across multiple cancer types, such as signature 1 (dominated by C>T mutation), which is shown to associate with age. Other signatures operate in only a subset of cancer types, such as signature 4 (dominated by C>A mutation), which is observed in lung, head and neck, and liver cancers. This signature is associated with tobacco smoking. More recently, the mutational signature catalogue has been updated and currently contains 49 single-base-substitution signatures. It should be noted that a recent study also extracts mutational signatures from other types of mutations than SNVs, including small indels. Overall, mutational signature analysis

is useful to associate observed mutational patterns with endogenous and exogenous factors that could influence carcinogenesis and cancer progression.

Another common alteration type is small indels. This refers to an addition or removal of a short sequence of nucleotides (Yang *et al.*, 2010). Indels can influence gene function by creating frameshift mutations and can often be the driver mutations in oncogenesis.

It should be noted that sometimes people use the word "somatic mutations" to refer specifically to SNVs and small indels but not SCNAs, while others also include SCNAs as somatic mutations. For simplicity throughout this thesis, I use the word "somatic mutations" when mentioning SNVs and small indels, but not SCNAs.

1.3.2.3 Copy-number alterations

SCNAs act on a large fraction of the genome, thus influencing tumourigenesis by affecting numerous genes simultaneously (Zack *et al.*, 2013; Harbers *et al.*, 2021). Any deviation from the basal 2n copy-number state of the genome can be considered a copy-number change. Thus, SCNAs include copy-number gain (>2n) and copy-number loss (1n or 0n) of any part of the genome. If the gains or losses occur on a chromosomal scale, it is referred to as aneuploidy. Interestingly, recent studies suggested that SCNAs, not somatic mutations, are a prognostic factor associated with recurrence and death (Hieronymus *et al.*, 2018; Smith & Sheltzer, 2018). SCNAs such as deletions can sometimes contribute to the loss of heterozygosity (LOH), the loss of one parental allele that might leave only the other aberrant allele. If LOH occurs at the tumour suppressor gene, it creates a cell that is more susceptible to malignant transformation (Ryland *et al.*, 2015).

Furthermore, whole-genome duplication (WGD), a doubling of the entire set of chromosomes, is common in human cancer. WGD creates genetically unstable cells that fuel tumour evolution (Bielski *et al.*, 2018). It has been suggested that WGD could buffer the deleterious effect of somatic alterations, particularly in the LOH regions (Lopez *et al.*, 2020). In addition, WGD is linked with a poor prognosis across cancer types (Bielski *et al.*, 2018). Therefore, WGD is an important event, and further investigation is needed to target this type of large-scale alteration in the cancer genome.

1.3.3 Age-associated differences in the cancer molecular landscape

In addition to an increase in cancer incidence and mortality with age, disparities between cancer in young and aged patients have also been observed (Van Herck *et al.*, 2021). For instance, breast cancer in younger patients tends to be more aggressive and associated with poorer survival (Anders *et al.*, 2008), while the prognosis is worse in older ovarian cancer patients (Maas *et al.*, 2005). Several studies have unravelled distinct molecular characteristics of tumours in relation to age in various cancer types, such as breast (Liao *et al.*, 2015; Kan *et al.*, 2018), prostate (Gerhauser *et al.*, 2018), and colorectal cancers (Lieu *et al.*, 2019). These analyses, however, focused on one cancer type and only a few molecular data types at a time. To bridge this gap, I performed an integrative analysis of the age-associated multi-omic landscape using a pan-cancer dataset from TCGA in **Chapter 3**. In this section, however, I summarise the major findings from previous cancer-specific studies and highlight important age-related genomic, transcriptomic, and epigenetic differences in cancers (**Table 1.1**). Content in this section is in part included in the accepted review article in *Trends in Cancer* entitled "Age-associated differences in the cancer molecular landscape".

1.3.3.1 Age-related genomic landscape in cancer

Increased age is associated with higher somatic mutations (single-nucleotide variants and small insertions/deletions) in most cancer types (Alexandrov *et al.*, 2015; Milholland *et al.*, 2015; Chalmers *et al.*, 2017). The spontaneous deamination of 5-methylcytosine to thymine (C>T) transitions, often referred to as the 'clock-like' mutational signature, dominates this age-related increase in mutation load. Furthermore, the mutational signature related to APOBEC cytidine deaminase activity also increases with age in prostate cancer (Gerhauser *et al.*, 2018). The observation of an age-related increase in somatic mutations could correspond to an accumulation of somatic mutations with age occurring before carcinogenesis, as recently reported in most non-cancerous human tissues (Martincorena *et al.*, 2018; Moore *et al.*, 2020; Kakiuchi & Ogawa, 2021).

Several cancer types display an age-associated mutational landscape in known cancer driver genes. In other words, some driver genes are mutated more often in younger or older individuals. A prominent example of this is a much higher frequency of mutations in *IDH1*, *ATRX*, and *TP53* in younger glioma patients. These mutations are associated with the IDH-

mutant subtype (Ceccarelli *et al.*, 2016). On the other hand, the IDH-wild-type subtype associated with copy-number losses of chromosome 10 (*PTEN*) and gains of chromosome 7 (*EGFR*) is higher in older glioma patients (Korber *et al.*, 2019).

Breast cancer is potentially the most well-characterized cancer in terms of age-associated subtyping. Oestrogen receptor-positive (ER^+) tumours are diagnosed more often in older individuals, while human epidermal growth factor receptor 2-positive ($HER2^+$) tumours are more common in younger patients. In another subtyping system, PAM50, older women were less diagnosed with the luminal B, HER2-enriched, and basal-like subtypes (Sweeney *et al.*, 2014). Regarding somatic mutations, higher *CDH1* mutations in older patients are observed (Liao *et al.*, 2015; Kan *et al.*, 2018). The mutations in *PI3KCA* also appear to increase with age (Mealey *et al.*, 2020). On the other hand, younger breast cancer patients are associated with higher *TP53* and *GATA3* mutations (Azim *et al.*, 2015; Kan *et al.*, 2018).

For other cancer types, it has been reported that *TP53* and *CTNNB1* mutations are more common in younger colorectal cancer patients, while *APC, KRAS,* and *BRAF V600* mutations are higher in older patients (Berg *et al.*, 2010; Willauer *et al.*, 2019). In contrast, *BRAF* mutations, especially *BRAF V600*, decrease with age in melanoma (Bauer *et al.*, 2011). In prostate cancer, structural variation breakpoints were highly enriched near gene regulatory regions such as active enhancers in the early-onset but not late-onset prostate cancer (Gerhauser *et al.*, 2018).

Overall, somatic mutations in cancer driver genes do not uniformly distribute across age. These age-related somatic mutations appear to be cancer-type specific. In addition, some mutations display opposite trends in different cancers. Typically, mutations in *BRAF V600* decline with age in melanoma and increase in colorectal cancer. Mutations in *BRAF V600* are an example of age-related mutations that are clinically actionable targets with multiple approved drugs (Tanda *et al.*, 2020). Thus, identifying age-related somatic mutations could have a profound impact in the clinical setting.

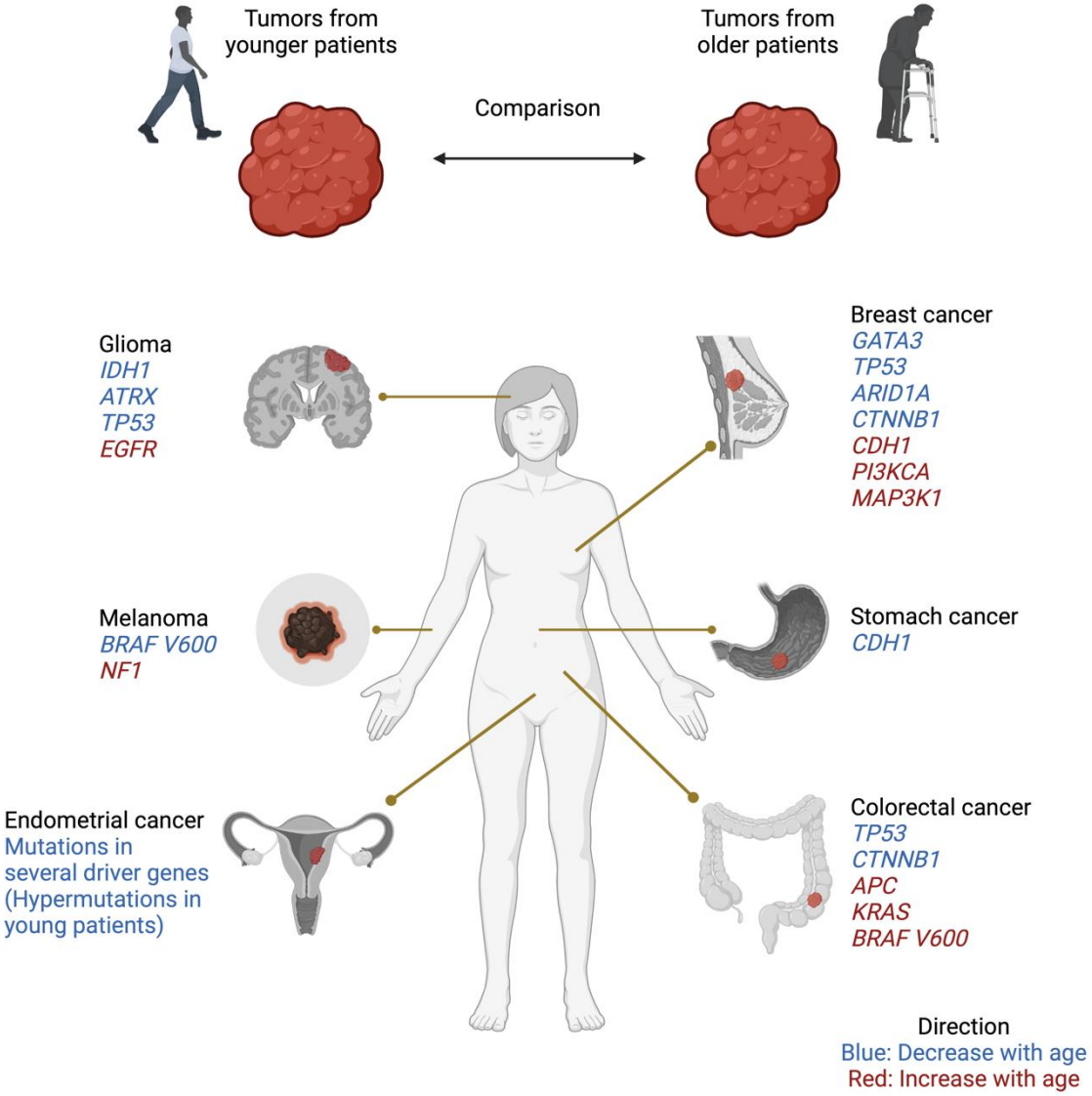


Figure 1.6. Age-associated differences in the cancer genome.

Example of cancer driver genes that display age-associated patterns in somatic mutations are shown in the figure. Driver genes in blue represent genes that are mutated more frequently in younger patients, while those in red denote genes that are mutated more often in older patients. Figure created with BioRender.com. This figure is a part of the accepted review article in *Trends in Cancer* entitled "Age-associated differences in the cancer molecular landscape".

1.3.3.2 Age-associated gene expression landscape in cancer

Several studies attempted to investigate age-related gene expression in cancer. In breast cancer, an early study reported a higher expression of cell cycle-related genes in tumours from younger than those from older ER^+ patients (Yau *et al.*, 2007). Recently, another work suggested that age-related differentially expressed genes in breast cancer could partly be controlled by age-related changes in oestrogen signalling (Osako *et al.*, 2020). A pan-cancer study reported that the amount of age-associated differentially expressed genes in cancer varies across tumour types (Wu *et al.*, 2019). The observed age-related gene expression changes in tumours are associated with numerous biological processes, such as extracellular matrix organization, metabolism, development, signalling pathways, and immune-related processes across various cancer types (Liao *et al.*, 2015; Kan *et al.*, 2018; Feulner *et al.*, 2019; Wu *et al.*, 2019). For instance, a recent study observed that ECM organization-related genes are upregulated with age in normal kidneys but are downregulated with age in clear cell renal cell carcinoma (ccRCC) (Feulner *et al.*, 2019). In addition, the expression of angiogenesis-related genes also decreases with age in ccRCC. However, Bozdag *et al.* found that angiogenesis-related genes are upregulated with age in glioblastoma (Bozdag *et al.*, 2013), highlighting cancer-type-specific gene expression differences with age.

As these results have been derived from bulk RNA-seq analyses, it is likely that they incorporate changes not only from cancer cells themselves but also from the ageing tissue microenvironment (Fane & Weeraratna, 2020). The ever-increasing data generated from scRNA-seq holds a great promise to resolving this issue. For example, a recent study revealed age-dependent alterations in cell proportions and gene expression, loss of tumour suppressor, and change in tissue microenvironment in mouse mammary gland (Li *et al.*, 2020a). Yet, to date, the comparison of gene expression between tumours as a function of patient's age using scRNA-seq is still lacking. It is possible that age-related alterations in the tissue microenvironment might provide different selective advantages for cancer cells containing distinct molecular alterations. This hypothesis is, however, waiting for experimental evidence.

Table 1.1: Examples of 'omics' studies of age-associated molecular alterations in cancer[a]

Cancer	Dataset	Molecular layer	Age group (Years old)	Key findings	Reference	PMID
Breast	TCGA	Genomics Transcriptomics	Young: ≤49 Middle-age: 46-69 Old: ≥70	- Somatic mutation and SCNA levels increased with age. - *GATA3* mutations were higher in younger patients. - Chr18p losses were higher in older patients, while Chr6q27 losses were found more often in younger patients.	Azim et al. (Azim *et al.*, 2015)	26467651
Breast	TCGA METABRIC	Genomics Transcriptomics DNA methylation Proteomics	Young: ≤45 Old: ≥55	- Gene expression differed between younger and older ER^+ breast cancers. - Several driver genes were differentially mutated between tumors from younger and older patients, such as higher mutations in *CDH1* in older patients. - Younger tumors are associated with elevated integrin/laminin and EGFR signaling.	Liao et al. (Liao *et al.*, 2015)	26251034
Breast	YWBC TCGA	Genomics	Young: ≤35 Old: ≥45	- In luminal A breast cancer, *PIK3CA* alterations were more common in older patients, whereas *GATA3* and *ARID1A* alterations were more common in younger patients.	Waks et al. (Waks *et al.*, 2022)	35101884
Breast	In-house dataset	Genomics Transcriptomics DNA methylation	Young: ≤45 Old: ≥70	- In ER^+ breast cancer, copy-number alterations were age-independent. - Gene expression differed between younger and older individuals, with cancer from younger groups expressing more cell cycle genes, and cancer from older groups expressing higher *HOX* genes and *ESR1*.	Yau et al. (Yau *et al.*, 2007)	17850661
Breast	TCGA	Genomics	Young: ≤40 Old: >40	- Older patients had a higher proportion of *PIK3CA*, *CDH1*, and *MAP3K1* mutations, while younger patients presented a higher *GATA3* and *CTNNB1* mutations. - Mutation load increased in older patients, with higher C>T mutations but lower C<A mutations with age.	Mealey et al. (Mealey *et al.*, 2020)	32164620

				- A class of tumors with a unique combination of mutational signatures was associated with a young age of onset.		
Breast	GEO (15 datasets)	Transcriptomics	Young: ≤40 Middle-age: 41-69 Old: ≥70	- Older age was associated with increased Luminal B and decreased basal-like subtypes. - Luminal B tumors in the oldest group were less aggressive than Luminal B tumors in the youngest group, based on different gene expression signatures.	Jézéquel et al. (Jezequel *et al.*, 2015)	26597277
Breast	SMC TCGA	Genomics Transcriptomics	Young: ≤40 Middle-age: 41-60 Old: >60	- A comparison between younger Asian breast cancer (SMC) and older Caucasian breast cancer (TCGA) cohorts revealed that the younger cohort had higher proportions of $HER2^+$ and Luminal B but lower Luminal A subtypes. - The germline pathogenic mutations (*BRCA1/BRCA2*) were significantly enriched in the younger cohort. - Somatic mutations in TP53 and *GATA3* were higher in the younger cohort, while mutations in *CDH1* were enriched in the older cohort. - The younger Asian cohort had a more immune-active microenvironment than the older Caucasian cohort.	Kan et al. (Kan *et al.*, 2018)	29713003
Breast	TCGA METABRIC Vancouver Big Series	Transcriptomics Protein expression (immunohistochemistry)	Young: ≤60 Old: >60	- A majority of age-related gene expression in cancer could be regulated by age-dependent estrogen signaling. - Age-related gene expression is mirrored by protein expression. - Gene encoding the chromatin modifier *EZH2* decline with age, while H3K27me3 levels increase with age.	Osako et al. (Osako *et al.*, 2020)	35121983
Colorectal	In-house dataset	Genomics	Young: <40 Old: ≥50	- Mutations in TP53 and *CTNNB1* decreased while mutations in *APC*, *KRAS*, and *BRAF* increased in older age group.	Lieu et al. (Lieu *et al.*, 2019)	31243121
Colorectal	MDACC AACR GENIE TCGA	Genomics	Young: <50 Old: ≥50	- Older patients had higher *BRAF V600* mutations. - The young adult patients (18-29 years old) had fewer *APC* mutations when compared with other early-onset patients (<50 years old).	Willauer et al. (Willauer *et al.*, 2019)	30854646

				- Early-onset patients (<50 years old) with inflammatory bowel disease were more likely to have *APC* mutations than early-onset patients without predisposing conditions.		
Colorectal	In house dataset	Genomics	Young: <50 Middle-age: 51-70 Old: >70	- *PIK3CA* mutations were not observed but TP53 mutations were more frequent in younger patients.	Berg et al. (Berg *et al.*, 2010)	21103049
Clear cell renal cell	TCGA CAGEKID	Transcriptomics	Age as a continuous variable	- Age-related gene expression in cancer and normal kidney tissues are different. - In normal tissues, upregulated genes with age were enriched for immune response and extracellular matrix organization, while downregulated genes were associated with metabolism and oxidation. - In cancer, age-related upregulated genes were enriched for metabolism and oxidation, while downregulated genes with age were enriched for extracellular matrix organization.	Feulner et al. (Feulner *et al.*, 2019)	30478010
Glioblastoma	TCGA REMBRANDT	Genomics Transcriptomics miRNAs DNA methylation	Young: ≤40 Old: ≥70 Age as a continuous variable	- The study discovered age-related signatures at the gene expression, epigenetic, and genomic levels. - Tumors from older patients presented increased Chr7 gains and Chr10 losses, but younger patients had Chr1q gains. - Older cancer is associated with the upregulation of angiogenesis-related genes. - In older patients, polycomb group protein target genes were hypermethylated.	Bozdag et al. (Bozdag *et al.*, 2013)	23658659
Low-grade glioma	TCGA GEO (3 datasets) CGGA	Transcriptomics DNA methylation	Young: ≤60 Old: >60	- The study identified differentially expressed genes with age in low-grade glioma. - Several upregulated genes with age were hypomethylated, such as *IGFBP2*, EMP3, *TIMP1*, and *SERPINE1*. High expressions of these genes were associated with a worse prognosis.	Wang et al. (Wang *et al.*, 2020)	32328202

Non-small-cell lung cancer	In-house dataset	Genomics	Age groups (<40, 40-49, 50-59, 60-69, and ≥70)	- The frequencies of mutations in *EGFR* and *ALK* were higher in younger patients. - The presence of these targetable alterations was associated with improved survival. - The youngest (< 40) and oldest groups (≥70) had the poorest survival compared with other age groups.	Sacher et al. (Sacher *et al.*, 2016)	26720421
Prostate	In-house dataset TCGA	Genomics Transcriptomics DNA methylation	Young: ≤55 Old: >55	- Somatic mutation and structural variation levels increased with age. - Recurrent gains of *ESRP1* were associated with high cell proliferation and poor outcomes in the younger cohort. - The mutational signature related to APOBEC3 increased with age. - Younger patients had a higher occurrence of the subgroup associated with a better prognosis.	Gerhauser et al. (Gerhauser *et al.*, 2018)	30537516
Prostate	In-house dataset TCGA	Transcriptomics miRNAs	Young: ≤45 Old: 71-74	- Differential expression analysis revealed higher inflammatory and immune responses in younger patients. - Younger patients showed a significant upregulation of CTLA4 and IDO1/TDO2 pathways.	Ding et al. (Ding *et al.*, 2016)	28027300
Pan-cancer	TCGA	Genomics Transcriptomics DNA methylation	Age as a continuous variable	- Mutation and SCNA levels increased with age. - Age-associated SCNAs and mutations were identified in several cancer driver genes across different cancer types, particularly gliomas and endometrial cancer. These age-related alterations in cancer driver genes could be associated with age-related cancer subtypes. - Age-related global transcriptomic changes were in part regulated by age-related DNA methylation changes. The age-associated genes were linked to numerous biological processes.	Chatsirisupachai et al. (Chatsirisupachai *et al.*, 2021)	33879792
Pan-cancer	TCGA PCAWG AACR GENIE	Genomics Transcriptomics	Age as a continuous variable	- Mutation and SCNA levels increased with age. - The study identified specific mutational signatures associated with age, such as tobacco use in tumors from younger patients and DNA damage repair signatures in older patients.	Li et al. (Li *et al.*, 2022)	35017538

				- Age-associated SCNAs and mutations were identified in several cancer driver genes across different cancer types. Some age-related mutations were age-dependent prognostic biomarkers, such as *SUFU* loss and *ATRX* in brain cancers.		
Pan-cancer	TCGA METABRIC	Genomics Transcriptomics DNA methylation	Young: First quartile Old: Fourth quartile Age as a continuous variable	- Using clinical data and gene expression analysis, this paper classified cancer into age-associated and non-age-associated tumor types. - Tumors in younger individuals were associated with accelerated molecular aging and senescence. - Aging was associated with tumor immune infiltration. - Multiple cancer driver genes showed age-associated mutational profiles.	Shah et al. (Shah *et al.*, 2021)	34879281
Pan-cancer	TCGA ICGC	Genomics Transcriptomics DNA methylation	Young: ≤50 Old: >50 Age as a continuous variable	- Mutation load, but not SCNAs, increased with age. - Age-associated SCNAs and mutations were identified in several cancer driver genes across different cancer types and cancer subtypes. These age-related mutations include several actionable genomic drivers with approved drugs. - Tumor immune landscape analysis revealed gene expression signatures and immune cell infiltration associated with young and old patients.	Lee et al. (Lee *et al.*, 2021)	34788626
Pan-cancer	TCGA AACR GENIE METABRIC	Genomics Transcriptomics DNA methylation	Age as a continuous variable	- Tumors from older patients were associated with favorable ICB markers, including increased mutation load, increased expression, and decreased methylation of immune checkpoint genes. - Tumors from older patients have decreased T cell receptor diversity, related to reduced ICB efficiency. - Deconvolution analysis revealed an age-related decrease in T cell abundance and an increase in macrophage abundance.	Erbe et al. (Erbe *et al.*, 2021)	34433020

Pan-cancer	TCGA	Transcriptomics	Young: <50 Old: >75	- The study used number of differentially expressed genes to classified cancer into the strong aging-effect and the weak aging-effect groups. - In the strong aging-effect group, older patients were associated with poor prognosis. - Differentially expressed genes with age were enriched in several pathways, such as epithelial-to-mesenchymal transition and KRAS signaling. - The expression of immune checkpoint genes mainly correlated with age.	Wu et al. (Wu *et al.*, 2019)	30230534
Stomach	TCGA	Transcriptomics miRNAs DNA methylation	Young: ≤49 Old: ≥70	- Cell cycle-related genes increased with age, while genes related to muscle process and adhesion decreased in older patients. - The age-related genes were potentially regulated by DNA methylation and miRNAs.	Kim et al. (Kim *et al.*, 2015)	25983541

[a]Abbreviations: AACR GENIE – American Association for Cancer Research Genomics Evidence Neoplasia Information Exchange; CAGEKID – Cancer Genomics of the Kidney; CGGA – Chinese Glioma Genome Atlas; GEO – Gene Expression Omnibus; ICGC – International Cancer Genome Consortium; MDACC – The University of Texas MD Anderson Cancer Center; METABRIC – Molecular Taxonomy of Breast Cancer International Consortium; PCAWG – The Pan-Cancer Analysis of Whole Genomes; REMBRANDT – The Repository of Molecular Brain Neoplasia Data; SMC – Samsung Medical Center in Korea; TCGA – The Cancer Genome Atlas; YMBC – Young Women's Breast Cancer Study

1.4 Ageing and development

1.4.1 The programmatic theory of ageing

One of the most influential theories in ageing research is antagonistic pleiotropy. Originally proposed by G.C. Williams in 1957, antagonistic pleiotropy argues that ageing evolves due to alleles that are beneficial for fitness in early life but cause harmful effects later (Williams, 1957). Natural selection fails to eliminate these alleles from the population because these deleterious effects do not appear until late life. Thus, unlike the damage-based theories, which propose that ageing arises from the stochastic accumulation of multiple forms of damage, antagonistic pleiotropy suggests that ageing has, to some degree, programmatic features and can be caused by the wild-type genome (Gems, 2022). However, it is unclear what biological mechanisms (biochemical, cellular, and physiological) could explain the antagonistic pleiotropy concept. More recently, the programmatic theory of ageing (the hyperfunction theory of ageing and the developmental theory of ageing) attempts to fill this gap by explaining that ageing-related changes are in part a product of regulated processes as part of developmental programmes (de Magalhaes, 2012; Gems, 2022).

Initially, development and ageing are considered to be different processes. However, several lines of evidence support the idea that ageing might originate from developmental programmes (Gems, 2022). First, the IIS and mTOR signalling pathways which can control ageing (section **1.1.3.2.5**) are essential in normal growth and development. Next, presbyopia, also known as age-related long sight, is caused by the continued, futile growth of the lens during adulthood. Another common example is the continuous growth of the prostate gland in men, which may also associate with an increased risk of prostate cancer in older men. Interestingly, the ratio between developmental time until sexual maturity and adult lifespan is approximately constant across mammalian species (Rayon & Briscoe, 2021; Gems, 2022). Thus, these examples suggest that the processes of ageing are in part originated from development. However, there was no switch to turn off developmental programmes in later life, potentially by the lack of natural selection force after the reproductive period. In turn, the developmental programmes persist into adulthood and may contribute to ageing. In other words, ageing can be viewed as an unintended outcome of development (de Magalhaes & Church, 2005).

1.4.2 Perturbations in development and the effects on ageing

If developmental processes influence ageing, then perturbations that are occurred during development should also affect lifespan and healthspan. Indeed, various maternal factors could influence the risks of developing age-related diseases, including cardiovascular diseases, obesity, diabetes, and cancer, affecting lifespan (Wadhwa *et al.*, 2009; Preston *et al.*, 2018). Malnutrition in utero is associated with a shorter lifespan in rats and mice (Chen *et al.*, 2009; Hanson *et al.*, 2016). In humans, maternal undernutrition results in lower birth weight of babies, which is associated with cardiovascular and metabolic disease risks later in life (Painter *et al.*, 2005; Preston *et al.*, 2018). Apart from nutrition, exposures to environmental toxicants such as cigarettes during gestation also correlate with late-life disease risks and reduction in lifespan (Power & Jefferis, 2002; Chen *et al.*, 2011; Rayfield & Plugge, 2017; Preston *et al.*, 2018).

Because of the profound links between development and ageing, recent studies investigated whether interventions that extend lifespan when applied in adolescence or later in life could also extend lifespan with early-life modulation. One recent study showed that short rapamycin treatment in a specific period early in life could increase longevity in mice and fruit flies (Aiello *et al.*, 2022). Interestingly, when the researchers treated rapamycin at later time points, the lifespan effect disappeared. Thus, lifespan can potentially be modulated only during development, indicating a link between early-life events and late-life effects on longevity. Likewise, another study treated rapamycin during the first 45 days of life in mice. The treated mice grew slower, but their median lifespan increased by 10%. In addition, the treated mice tend to preserve better health as measured by several physiological indexes and epigenetic clock. Overall, these studies suggest a profound link between developmental pace and longevity (Shindyapina *et al.*, 2022).

1.4.3 Transcriptomic and epigenetic links between development and ageing

Using large-scale transcriptomic and epigenetic assays, several studies attempted to examine the relationship between development and ageing gene expression and the regulation of these genes. An early work studied gene expression changes in mouse organs (liver, lung, and kidney) from early to late life (Lui *et al.*, 2010). The results revealed that many changes in gene expression occurring during ageing originate during development, including a persistent

decline in cell-cycle-related genes. Thus, this study argues that the genetic programme that limits juvenile growth persists into adulthood and supports the programmatic theory of ageing (Lui *et al.*, 2010; de Magalhaes, 2012).

Another study from Somel *et al.* investigated microRNA (miRNAs), mRNA, and protein expression in postnatal development and ageing in human and macaque brains (Somel *et al.*, 2010). They found that the majority of mRNA expression changes originated in development. Some ageing gene expressions represent reversals of developmental trajectories, while other genes show extensions of developmental patterns. For instance, DNA repair genes decrease during development but increase during ageing. This group of genes represents a reversal pattern, possibly as a response to damage. On the other hand, neuronal-related genes rise during development but decline during ageing, highlighting a reversal pattern that might be a consequence of damage. Indeed, there are also groups of genes whose expression showed a continuation from development to ageing. Genes related to neural development, axon guidance, cell communication, and cell-cell adhesion showed a continuation trend by decreasing from early childhood to late life. These groups of genes could represent an uncontrolled or "runaway" continuation of developmental programmes. Furthermore, this study also identified miRNAs and transcription factors that control both developmental and ageing processes, suggesting potential regulatory mechanisms for the development-ageing genes. Together, this study provides possible links between development and ageing at transcriptomic and epigenetic levels, supporting the programmatic theory of ageing.

Interestingly, following works from Somel and colleagues reported that a set of genes involving neuronal and synaptic functions upregulate during development but downregulate during ageing. They argue that this up-down reversal pattern may be associated with ageing-related loss of cellular identity in neurons (Donertas *et al.*, 2017). A comparison between tissues in mice further strengthens this idea by revealing that inter-tissue gene expression diverges during development and converges during ageing (Izgi *et al.*, 2022). The divergent-convergence pattern was prevalent among tissue-specific genes and associated with loss of tissue identity. Thus, emerging studies reveal a close link between gene expression changes across the lifespan, from development to ageing. However, further studies are required to better characterise these gene sets.

The relationship between development and ageing is not limited to the transcriptome level. Notably, methylation sites of DNA methylation clocks (epigenetic clocks) are associated with developmental-related genes. Furthermore, DNA methylation clocks run the fastest during development, and much slower during adulthood (Simpkin *et al.*, 2016; Bell *et al.*, 2019), suggesting that the ageing process could be a part of development (Raj & Horvath, 2020). Another study in *C. elegans* identified chromatin accessibility changes during development and ageing. This work further showed that binding sites of some transcription factors were enriched in dynamic accessible regions in both development and ageing (Janes *et al.*, 2018).

Taken together, the relationship between development and ageing is found in different layers of information, from gene expression to epigenetics. However, the studies mentioned above are largely based on postnatal development. It is still unclear if gene expression and epigenetic changes during prenatal development also relate to those altered during ageing. In **Chapter 4**, I employed RNA-seq and chromatin immunoprecipitation (ChIP)-seq datasets generated from mice to investigate the relationship between changes in gene expression and histone modification during prenatal development and ageing.

1.5 Thesis aims and overview

The main objectives of this thesis are to employ publicly available datasets to better understand the relationship between human ageing and cancer and the links between mammalian development and ageing. Here, I set out five specific aims as follows:

1. To investigate the relationship between human ageing and cancer in a tissue-specific manner.
2. To derive cellular senescence signatures, genes that are consistently overexpressed or underexpressed in senescent cells across several datasets, and to compare the cellular senescence signatures with ageing and cancer genes.
3. To examine the molecular landscape in cancer according to patient's age and identify age-associated cancer driver genes, genes that are mutated more or less frequently in younger or older patients.
4. To study the differences in gene expression and DNA methylation patterns in tumours derived from patients of different ages.
5. To investigate the relationship between dynamic gene expression and histone modification changes during development and ageing.

To achieve these aims, the studies in this thesis leverage publicly available large-scale omics datasets to unravel links between ageing, cancer, and development. In **Chapter 2**, I used gene expression datasets from GTEx, TCGA, and others to achieve aims 1-2 by exploring the relationship between transcriptomic changes in ageing, cancer, and cellular senescence. Next, I asked whether tumours from patients of different ages show similar or different molecular aberrations? In **Chapter 3**, I took advantage of multi-omic datasets from TCGA, covering 33 cancer types and ~10,000 tumours, to comprehensively characterise the age-associated genomic, transcriptomic, and epigenetic landscape across cancer types, answering aims 3-4. Finally, I analysed gene expression and histone modification from The Encyclopedia of DNA Elements (ENCODE) project and another dataset in **Chapter 4** to study the molecular links between mammalian prenatal development and ageing, thus accomplishing aim 5.

Chapter 2

A human tissue-specific transcriptomic analysis of the relationship between ageing, cancer, and cellular senescence

2.1 Introduction

Gene expression is a fundamental mechanism of a cell to use genetic information stored in DNA to produce an observable phenotype of a gene via RNA synthesis. Gene expression is highly dynamics across biological conditions. And changes in transcriptomes are associated with numerous biological processes, including ageing (de Magalhaes *et al.*, 2009; Palmer *et al.*, 2021) and diseases such as cancer (Cieslik & Chinnaiyan, 2018). During the past two decades, transcriptome-wide analysis, initially by microarray and then RNA-seq, has been used to study most, if not all, aspects in biological sciences (Wang *et al.*, 2009; Stark *et al.*, 2019). On top of that, large-scale international consortium efforts such as The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research *et al.*, 2013b) and The Genotype-Tissue Expression (GTEx) (Consortium, 2013) have generated unprecedented sequencing data to better understand the biology of cancer and tissue-specific gene expression and regulation, respectively. Due to the ever-growing amounts of transcriptomic data, it is currently possible to compare changes in gene expression between different biological processes, including those changes in ageing, cancer, and cellular senescence.

While ageing has long been recognised as a primary risk factor for cancer, the relationship between these two processes has been poorly understood (de Magalhaes, 2013). Gene expression changes in ageing may inform us about the biological processes changing with age in each tissue. Furthermore, by comparing gene expression changes between ageing and cancer, a better understanding on how tissue-specific ageing relates to cancer could be obtained. A recent study has reported that transcriptomic changes during ageing tend to shift away from cancer and move toward degenerative disease signatures (Aramillo Irizar *et al.*, 2018). However, this study focuses on the overall pattern and include only four human tissues, precluding the tissue-specificity of the relationship between ageing and cancer. This chapter investigated gene expression changes during ageing from 9 human tissues from GTEx. I next

compared transcriptomic changes in ageing with those changes between cancer and matched normal tissue from 9 corresponding TCGA cancer types.

Cellular senescence, a state of irreversible growth arrest of cells, increases in ageing tissues. While cellular senescence has been regarded as an anti-tumour mechanism, evidence has suggested that senescent cells could also promote cancer via their secretory phenotypes (Campisi, 2013). In the next part of this chapter, I performed a meta-analysis of 20 cellular senescence microarray gene expression datasets to derive cellular senescence signature genes, genes that are consistently overexpressed or underexpressed in senescent cells across several datasets. I used these senescence signatures to examine the relationship between transcriptomic changes in cellular senescence and ageing and senescence and cancer. Part of results in this chapter have been published in *Aging Cell* (Chatsirisupachai *et al.*, 2019).

2.2 Materials and Methods

2.2.1 Linear regression to identify tissue-specific genes differentially expressed with age (age-DEGs) from GTEx

The RNA-Seq based gene expression data of non-cancerous tissues (v7, January 2015 release) were downloaded from the GTEx portal (https://gtexportal.org) (Consortium, 2015). RNA-Seq was performed using the Illumina TrueSeq library construction protocol. Reads were aligned to the human reference genome hg19/GRCh37 based on the GENCODE v19 and were processed as described in https://gtexportal.org/home/documentationPage. Out of 30 tissues provided by GTEx, four tissues (bladder, cervix uteri, fallopian tube, and kidney) with low samples (11, 11, 7, and 45, respectively) were excluded from the analysis. For each tissue, differentially expressed genes with age were identified using the following linear regression model:

$$Y_{ij} = \alpha Age_i + \beta Sex_i + \gamma Death_i + \varepsilon_{ij}$$

Where Y_{ij} is the expression level of gene j in sample i, Age_i denotes the age of sample i. Sex_i denotes the sex of sample i, $Death_i$ denotes the death classification of sample i based on the 4-point Hardy scale (Ferreira *et al.*, 2018), and ε_{ij} represents the error term. The dataset downloaded from the GTEx portal did not provide the actual age of each sample; the age ranges, i.e. 20-29, 30-39, 40-49, 50-59, 60-69 and 70-79, were provided instead. The age of each sample was then approximated to 25, 35, 45, 55, 65 and 75, respectively. Because our main aim was to compare the age-DEGs with cancer-DEGs from TCGA, and there was a large difference between the number of the total genes in GTEx (56,202) and TCGA (20,532), only on protein-coding genes were focused. Protein-coding genes were identified using the R package *biomaRt* (version 2.36.1) (Durinck *et al.*, 2009), based on Ensembl release 92 (April 2018). In total, 18,851 protein-coding genes were left after removing the non-coding genes, and these genes were used in the subsequent analyses. To remove lowly expressed genes, genes with an expression of less than one count per million (cpm) in more than 30 per cent of samples were excluded. Raw read counts were normalised using TMM normalisation and were voom transformed to remove heteroscedasticity from the count data. The linear model for each gene was generated using the *limma* package in R (version 3.36.5) (Ritchie *et al.*, 2015). Genes were considered to be significantly differentially expressed genes with age if the empirical Bayes

moderated t-statistics and their associated adjust *p*-value (Benjamini-Hochberg method) < 0.05 and absolute fold change across 50 years of age (from 25 to 75 years old) > 1.5. The genes in GTEx were in ensemble id, thus, they were converted into Entrez id format using *biomaRt*.

Nine tissues were selected for further analysis, including breast, colon, oesophagus, liver, lung, prostate, stomach, thyroid and uterus. This selection was based on the criteria that 1) each of these tissues had a matched TCGA project from the same tissue of origin and 2) the number of solid normal samples in those TCGA projects were more than 10.

2.2.2 Differential expression analysis to identify differentially expressed genes in cancer (cancer-DEGs)

The TCGA level 3 RNA-Seq based gene expression data for 10 TCGA projects used in the analysis (BRCA (Cancer Genome Atlas, 2012b), COAD (Cancer Genome Atlas, 2012a), ESCA (Cancer Genome Atlas Research *et al.*, 2017), GBM (Brennan *et al.*, 2013), LIHC (Cancer Genome Atlas Research Network. Electronic address & Cancer Genome Atlas Research, 2017), LUAD (Cancer Genome Atlas Research, 2014b), PRAD (Cancer Genome Atlas Research, 2015), STAD (Cancer Genome Atlas Research, 2014a), THCA (Cancer Genome Atlas Research, 2014c), and UCEC (Cancer Genome Atlas Research *et al.*, 2013a)) were downloaded from FireBrowse (http://firebrowse.org) in June 2019. RNA sequencing was performed on the Illumina Hi-Seq platform and aligned to the human reference genome GRCh37. The data was then processed using the RNASeqV2 pipeline of TCGA, which provided the RSEM expected counts. Protein-coding genes were identified using the *biomaRt* R package (version 2.36.1). In total, 18,548 protein-coding genes were left after removing the non-coding genes, and these genes were used in subsequent analyses. To remove lowly expressed genes, genes with an expression of less than one count per million (cpm) in more than 30 per cent of samples were excluded. The *limma* package in R was used to identify genes significantly differentially expressed between cancer and normal samples. The moderated t statistic *p*-value after correction by Benjamini-Hochberg method < 0.01 with an absolute fold-change > 2 was used to determine the statistical significance.

2.2.3 Meta-analysis to identify cellular senescence signature genes

To identify cellular senescence signature genes, a meta-analysis of 20 replicative senescence GEO datasets was performed. The datasets included in the analysis were gathered by Miss Susana Ferreira and are shown in **Table 2.3**. Differentially expressed genes were identified seperately in each dataset. The probe annotation for each dataset differed depending on the platform. In general, the steps are 1) Download the dataset from GEO, 2) Probe annotation to match probe ID with Entrez ID using either the dataset's own GPL platform files, R packages (*AnnotationDbi* version 1.44.0 and *org.Hs.eg.db* version 3.7.0), or bioDBnet (https://biodbnet-abcc.ncifcrf.gov/db/db2db.php) (Mudunuri *et al.*, 2009) 3) Remove probes which do not match to any Entrez ID 4) Remove probes which match to multiple Entrez ID 5) Average intensity of probes which match to the same Entrez ID. The differential expression analysis was performed using limma. Genes with a *p*-value below 0.05 and absolute fold change > 1.5 were considered putatively differentially expressed genes. For the datasets with only one non-senescent sample and one senescent sample, only fold change cut-off was used to determine differentially expressed genes. There are two platforms (GPL2937 and GPL2947) in the dataset GSE3460, so each platform was separately analysed and they were treated as two different datasets in the meta-analysis. After that, the results from all datasets were then combined using a binomial distribution and the false discovery rate set at $Q < 0.05$ using the same method as (de Magalhaes *et al.*, 2009).

2.2.4 Protein-protein interaction (PPI) network analysis

The PPI network was built based on the physical multi-validated protein interactions data from the BioGrid database (Version 4.4.201) (Oughtred *et al.*, 2021). Cellular senescence signature genes, together with their first-order interactions, were extracted. The network was constructed and visualised in Gephi (Version 0.9.2) (Bastian *et al.*, 2009). Nodes were coloured as follows: overexpressed signature – light brown, underexpressed signature – light blue, and non-senescence signature – grey. Network topology analysis was performed in Gephi.

2.2.5 Fold change with age in normal tissues of cancer-DEGs and cellular senescence signatures

Fold change with age of the cancer-DEGs and cellular senescence signature genes were examined. After obtaining the cancer-DEGs from TCGA, the fold change with age in normal tissues in GTEx of these cancer-DEGs was identified. Although not every cancer-DEG was found in the GTEx data from the corresponding tissue because lowly expressed genes had been filtered out before performing linear regression, most of them retained. For each tissue, the fold change with age in normal tissue of up-regulated genes in cancer and down-regulated genes in cancer were compared using a two-sided Wilcoxon rank-sum test. Similarly, the fold change with age in normal tissues in GTEx of cellular senescence signature genes, as well as the fold change in cancer from TCGA of cellular senescence signature genes were also determined.

2.2.6 Overlap analysis

The overlap analyses were performed using the *GeneOverlap* package in R (version 1.16.0). In each experiment, background was set differently depending on where the genes come from. For the overlap between age-DEGs from GTEx and cancer-DEGs from TCGA, all protein coding genes in GTEx data (18,851 genes) were used as a background. For the overlap between age-DEGs and cellular senescence signature genes, all protein coding genes in combined datasets used for the meta-analysis (18,878 genes) were used as a background. For the overlap between cancer-DEGs and cellular senescence signature genes, all protein coding genes in combined datasets used for the meta-analysis (18,878 genes). The overlap was considered significant if an adjusted p-value < 0.05 (Fisher's exact test followed by Benjamini-Hochberg correction).

2.2.7 Functional enrichment analysis

Gene ontology (GO) enrichment analysis was performed using the *clusterProfiler* package in R (version 3.8.1) (Yu *et al.*, 2012; Wu *et al.*, 2021). A GO term was considered to be an enriched term if an adjusted p-value < 0.1 (Benjamini-Hochberg correction). The union set of age-DEGs and cancer-DEGs in each of the four conditions for each tissue was employed as a background for GO enrichment test of the overlapping gene set in that tissue. For example, the union set of genes up-regulated with age in colon and genes up-regulated in COAD was used

as a background for GO enrichment analysis of overlapping genes between genes up-regulated with age in colon and genes up-regulated in COAD. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed on the underexpressed and overexpressed cellular senescence signatures. A pathway was enriched if an adjusted *p*-value < 0.05 (Benjamini-Hochberg correction). All protein coding genes in combined datasets used for the meta-analysis (18,878 genes) were employed as a background.

2.2.8 Validating the results using GTEx and TCGA data from recount2

Because GTEx and TCGA had processed the data differently, I confirmed that this has not affected the results of the comparison between age-DEGs and cancer-DEGs by using data from recount2 (Collado-Torres *et al.*, 2017). Recount2 is a project to standardize the pipeline of analysis and provide the ready-to-analyze data, thus TCGA and GTEx (version 6) data have been processed using the same pipeline and they contain the same number of genes (58037 genes in total, including 19732 protein-coding genes). The analyses following the same methods as describe above were performed using this recount2 data and confirmed that age-DEGs and cancer-DEGs change in the same direction in thyroid and uterus, while change toward the opposite direction in the other tissues. I, however, keep the results analysed from the GTEx database (version 7) and TCGA downloaded from FireBrowse, because GTEx (version 6) contain less samples, particularly in the oldest age group (70 – 79 years old).

2.3 Results

2.3.1 Human tissue-specific differentially expressed genes with age

I first identified tissue-specific age-DEGs from 26 GTEx tissues using linear regression analysis. The numbers of samples included in the analysis are shown in **Table 2.1**. To gain an overview of gene expression changes with age across tissues, I considered the Pearson correlation of expression fold changes with age across tissues, using only genes expressed in all tissues. Although the correlation coefficients were not high in general (-0.25 – 0.60, median = 0.13), I found that tissues that work together in the same systems or share similar germ layer origins tended to cluster together (**Figure 2.1**). For instance, tissues from digestive tracts, including the stomach, small intestine, and colon, were clustered together. Similarly, the ovary and uterus, organs from the female reproductive system, were grouped. Heart and muscle consisted of muscle fibres, were in the same cluster, together with blood vessel, oesophagus, and adipose tissue. These tissues were developed from a mesoderm layer. However, it is still unclear why some tissues were correlated in terms of age-related gene expression changes. For example, blood was in the same cluster as the brain and pituitary. One possibility is that immune cell infiltration occurs in the ageing brain (Gemechu & Bentivoglio, 2012), driving to some degree a similar age-related change in blood and brain transcriptomes. Overall, tissues that were more similar in origin and function tended to share ageing gene expression profiles.

The numbers of significant age-DEGs from the regression analysis (adj. *p*-value < 0.05 and absolute fold change > 1.5; moderated t-test) varied considerably between different tissues from 1,369 in the brain to 1 in the small intestine (**Figure 2.2**). I found that age-DEGs were mostly tissue-specific, with 2419 (76%) of 3183 unique downregulated genes and 2225 (62%) of 3585 unique upregulated genes were identified in only 1 tissue (**Figure 2.3**). This tissue-specificity of age-DEGs is consistent with a previous report (Yang *et al.*, 2015).

Table 2.1: Number of GTEx and TCGA samples included in this study

GTEx Tissues		TCGA Projects			
Tissue Name	Cases	Disease Name	Project	Samples	
				Primary Solid Tumour	Solid Tissue Normal
Breast	290	Breast invasive carcinoma	BRCA	1093	112
Colon	506	Colon adenocarcinoma	COAD	285	41
Esophagus	1018	Esophageal carcinoma	ESCA	184	11
Liver	174	Liver hepatocellular carcinoma	LIHC	371	50
Lung	425	Lung adenocarcinoma	LUAD	515	59
Prostate	152	Prostate adenocarcinoma	PRAD	497	52
Stomach	261	Stomach adenocarcinoma	STAD	415	35
Thyroid	444	Thyroid carcinoma	THCA	501	59
Uterus	110	Uterine Corpus Endometrial Carcinoma	UCEC	186	24
Adipose tissue	786	No match TCGA project with more than 10 solid tissue normal			
Adrenal gland	190				
Blood	515				
Blood vessel	901				
Heart	599				
Muscle	552				
Nerve	406				
Ovary	131				
Pancreas	248				
Pituitary	183				
Salivary gland	97				
Skin	1181				
Small intestine	137				
Spleen	162				
Testis	259				
Vagina	113				

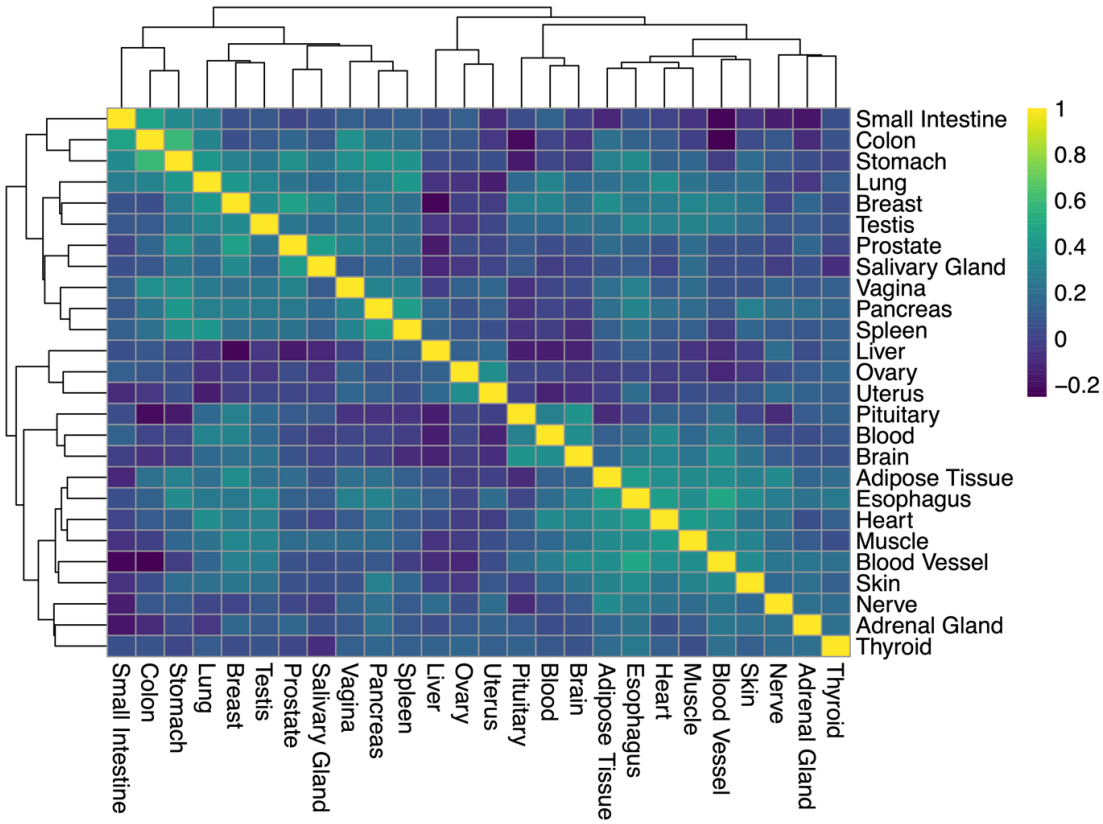


Figure 2.1. Correlation of the age-related gene expression changes across 26 GTEx tissues. Pairwise Pearson's correlation of the gene expression fold changes with age between pairs of tissues was calculated and plotted.

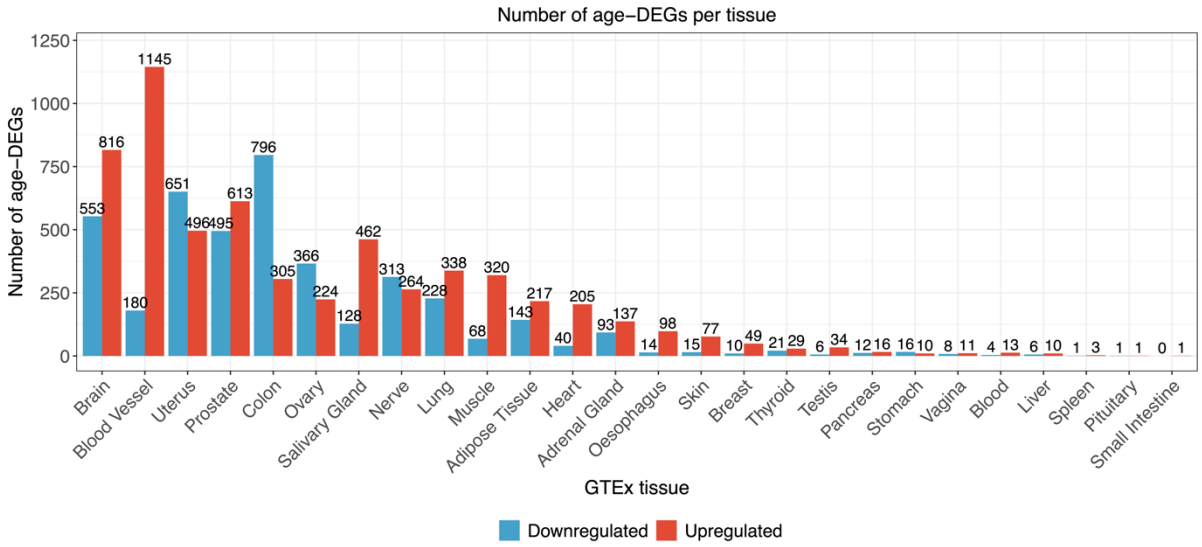


Figure 2.2. The number of age-DEGs per GTEx tissue. Linear regression analysis was used to identify age-DEGs. The bar chart ranges from tissue with the highest number of age-DEGs to tissue with the lowest number of age-DEGs. The number of downregulated genes was shown in blue bars, whilst the number of upregulated genes was shown in red bars.

Only one downregulated gene (*NEFH*) was commonly identified in 9 tissues. The *NEFH* gene encodes the heavy neurofilament protein, a cytoskeletal component of the axon. Mutations in this gene have been linked with amyotrophic lateral sclerosis (ALS) (Simpson & Al-Chalabi, 2006). The most common upregulated gene with age was *EDA2R*, found in 17 out of 26 tissues. The *EDA2R* gene encodes a protein in TNFR (tumour necrosis factor receptor) superfamily. By using the same GTEx data, de Vries et al. has reported *EDA2R* as a strong candidate gene for lung ageing (de Vries *et al.*, 2017; Jeong *et al.*, 2020). Moreover, *EDA2R* has been shown to be a target of p53 tumour suppressor protein (Brosh *et al.*, 2010). Furthermore, recent work has shown that the upregulation of EDAR2 is accompanied by an increase of its ligand in the bloodstream. An increased *EDA2R* expression was also found in ageing gastrocnemius in mice and rats, highlighting the conservation across species. Finally, *EDA2R* is among the top upregulated genes in the mouse model of the Hutchinson Gilford progeria syndrome (Maria Chiara *et al.*, 2021). *LMO3* is the second most common upregulated age-DEGs, shared among 14 tissues. LMO3, a member of the LIM domain only protein family, plays a role in transcriptional regulation and has been associated with carcinogenesis (Liu *et al.*, 2015). It has previously been reported to be upregulated with age in skin and adipose tissues (Glass *et al.*, 2013), our data also confirmed this notion.

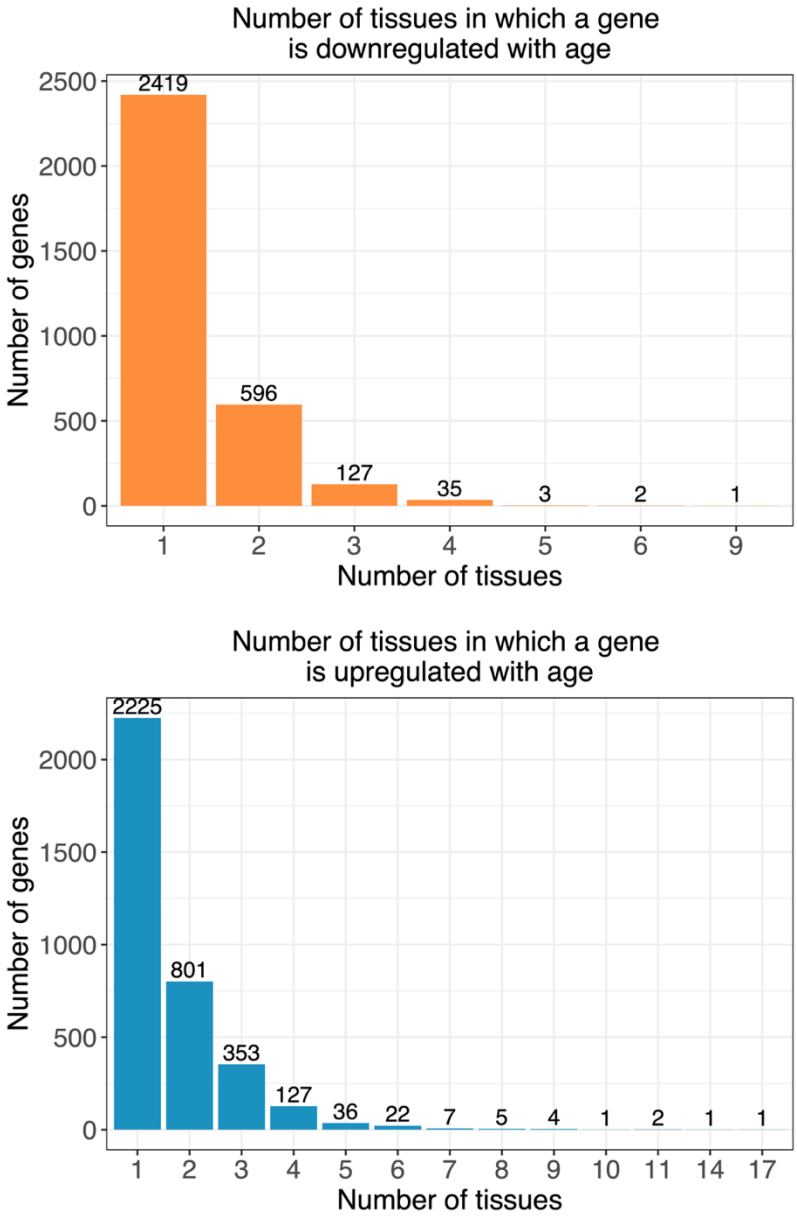


Figure 2.3. The number of tissues in which an age-DEG is downregulated or upregulated in.

(Top) The number of tissues in which a gene is significantly downregulated with age. **(Bottom)** Number of tissues in which a gene is significantly upregulated with age.

I next overlapped our gene lists with existing ageing-related gene lists. I found that 107 out of all 5,580 age-DEGs (upregulated or downregulated genes with age in at least one tissue) were included in the GenAge database of human ageing-related genes (Tacutu *et al.*, 2018) (*p*-value = 0.0116, odds ratio (OR) = 1.33, **Figure 2.4A**). I retrieved ageing gene signatures from a recent meta-analysis from Palmer *et al.*(Palmer et al., 2021), consisting of 449 overexpressed signature genes and 162 underexpressed signature genes. I found a significant overlap between

upregulated age-DEGs in at least one tissue and overexpressed signatures (**Figure 2.4B**, *p*-value = 2.63 x 10^{-52}, OR = 4.59). Similarly, downregulated age-DEGs in at least one tissue were overlapped with underexpressed ageing gene signatures (**Figure 2.4C**, *p*-value = 0.00142, OR = 1.78). I further check the overlap between age-DEGs in at least one tissue and Palmer's ageing gene signatures that change in the opposite direction. Notably, upregulated age-DEGs were not overlapped with underexpressed signatures (**Figure 2.4D**, *p*-value = 0.693, OR = 0.92). However, the overlap between downregulated age-DEGs and overexpressed signatures was significant (**Figure 2.4E**, *p*-value = 0.00357, OR = 1.39), albeit the effect size smaller than the overlap between upregulated age-DEGs and overexpressed signatures (OR = 4.59, **Figure 2.4B**). Taken together, these overlap analyses between our age-DEGs and known ageing gene lists reinforce the reliability of our age-DEG lists.

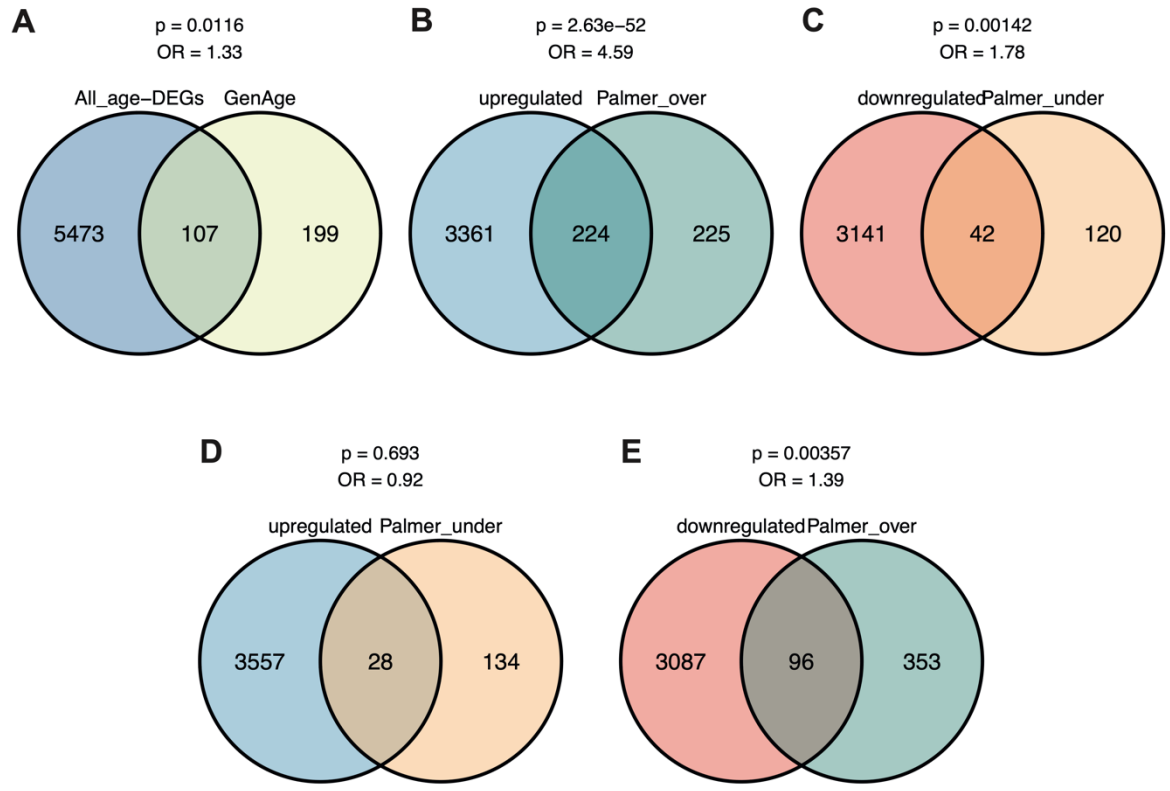


Figure 2.4. Overlap analysis between age-DEGs and known ageing gene sets.

Fisher's exact test was performed to investigate the overlap between **(A)** all age-DEGs (downregulated and upregulated genes) and ageing-related genes from the GenAge database, **(B)** upregulated age-DEGs and overexpressed ageing signatures, **(C)** downregulated age-DEGs and underexpressed ageing signatures, **(D)** upregulated age-DEGs and underexpressed ageing signatures, and **(E)** downregulated age-DEGs and overexpressed ageing signatures.

To determine biological processes underpinning tissue-specific age-related gene expression changes, I performed gene ontology (GO) enrichment analysis on upregulated and downregulated genes with age. Age-DEGs from different tissues were enriched in a wide variety of GO terms (**Table 2.2**). It should be noted that not every case has an enriched term, possibly due to the low number of genes differentially expressed with age in some tissues. Here, I presented some examples of enriched terms. The most prominent enriched term in upregulated genes was immune response processes, enriched in several tissues such as adipose tissue, brain, and prostate. Immune response genes have been found to be overexpressed with age from the meta-analysis of several ageing transcriptome datasets (de Magalhaes *et al.*, 2009; Palmer *et al.*, 2021). This upregulation in immune-related genes with age could represent inflammaging, persistent stimulation of the innate immune system that could contribute to age-related diseases (Ferrucci & Fabbri, 2018; Franceschi *et al.*, 2018b). In contrast, the immune response was enriched in genes downregulated with age in the colon. This downregulation of immune-related genes could partly be explained by the compromised function with age of the mucosal immune system in the intestine (Mabbott *et al.*, 2015). Terms associated with ECM were also enriched in upregulated age-DEGs in the heart, lung, and muscle and downregulated age-DEGs in adrenal gland and uterus. Changes in ECM integrity during ageing could create a pervasive environment for cancer progression (Fane & Weeraratna, 2020). In many tissues, enriched terms related to age-DEGs were developmental processes and cell cycle regulation. These terms could reflect alterations in stem cell population and gene expression during ageing (Lopez-Otin *et al.*, 2013). Another possible interpretation is that these age-DEGs represent age-associated transcriptomic alterations exhibiting developmental mechanisms that continue throughout adulthood (de Magalhaes, 2012). Overall, age-DEGs were tissue-specific and were associated with numerous biological processes related to tissue pathology and age-related diseases.

Table 2.2: GO enrichment analysis results of tissue-specific age-DEGs

Tissue	Downregulated	Upregulated
Adipose tissue	Metabolic process	Immune response
Adrenal gland	Extracellular matrix organisation	Immune response
Blood vessel	Developmental process	Immune response
Blood	JNK cascade, B signalling	Neurotransmitter receptor activity
Brain	Synaptic process	Immune response
Breast	-	Cell-substrate adhesion
Colon	T cell activation	Muscle development
Oesophagus	Transport, muscle contraction	-
Heart	Ribonucleotide catabolic process	Extracellular matrix organisation
Liver	-	-
Lung	Cell cycle, DNA replication	Extracellular matrix organisation
Muscle	Secondary metabolic process	Extracellular matrix organisation
Nerve	Metabolic process, vasoconstriction	Synaptic process
Ovary	Cell proliferation, developmental process	Muscle filament sliding
Pancreas	-	-
Pituitary	Peptide and lipoprotein metabolic process	Protein trimerization, ion transport
Prostate	Amino acid metabolism and transport	Immune response
Salivary gland	-	Developmental process, cell proliferation
Skin	Branched-chain amino acid transport	Synaptic process, muscle filament sliding
Small intestine	-	Developmental process
Spleen	Cytoskeleton	-
Stomach	-	CDK activity, cellular senescence, p53
Testis	-	BMP signalling pathway
Thyroid	-	-
Uterus	Extracellular matrix organisation	Biosynthetic process, cell cycle
Vagina	Myeloid leukocyte differentiation	-

2.3.2 Differentially expressed genes between primary tumours and matched normal tissues

I analysed TCGA level 3 normalised RNA-Seq data to identify differentially expressed genes between primary tumours and matched normal tissues. Nine TCGA projects used in our analysis were BRCA, COAD, ESCA, LIHC, LUAD, PRAD, STAD, THCA, and UCEC. The number of primary tumours and normal tissue included in each project was summarised in **Table 2.1**. Using an empirical Bayes moderated t-test in the *limma* R package, the significant down-regulated and upregulated genes (adj. *p*-value < 0.01, absolute fold change > 2) in tumour compared to normal samples were determined as shown in **Figure 2.5**.

Overall, numbers of differentially expressed genes between cancer and matched normal tissues (cancer-DEGs) were relatively higher when compared with age-DEGs, ranging from about 1,700 cancer-DEGs (combining upregulated and downregulated genes) in PRAD to nearly 3,900 genes in UCEC. In addition, cancer-DEGs were generally shared among cancer types, with 3,558 (57.40%) of 6,199 unique downregulated genes and 2,040 (47.90%) of 4,259 unique upregulated genes were shared between two or more cancer types (**Figure 2.6**). In downregulated and upregulated genes, 32 and 36 genes were shared between all 9 TCGA projects, respectively. Gene ontology (GO) enrichment analysis revealed that the function of 36 common up-regulated genes in all TCGA projects was mainly related to the cell cycle (adjust *p*-value by Benjamini-Hochberg < 0.05). However, there were no significantly enriched terms in 32 shared down-regulated genes in all TCGA projects.

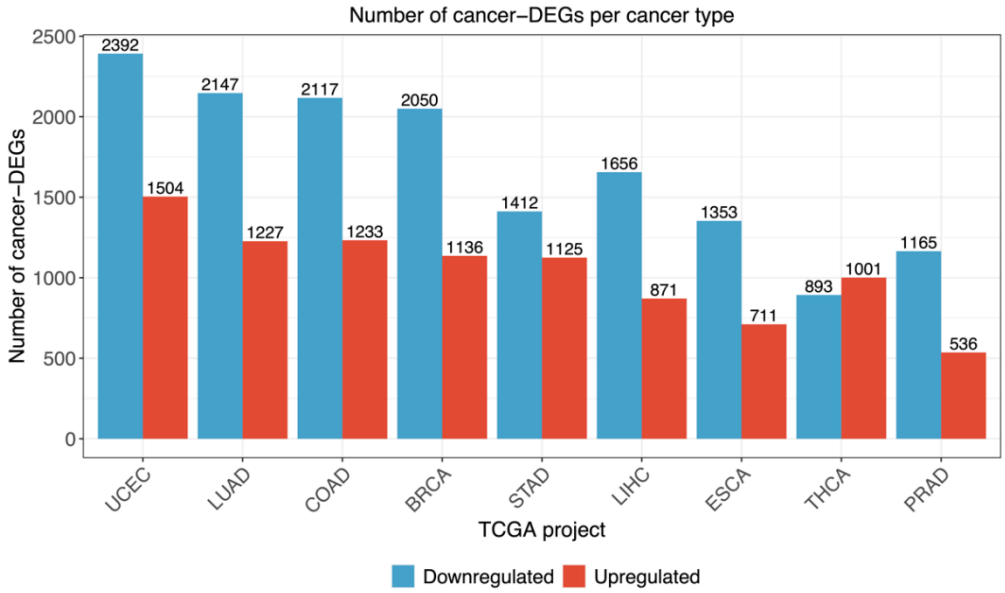


Figure 2.5. The number of cancer-DEGs per TCGA cancer type.

Differential expression analysis was used to identify cancer-DEGs, differentially expressed genes between primary tumours and matched normal tissues. The bar chart ranges from cancer type with the highest number of cancer-DEGs to tissue with the lowest number of cancer-DEGs. The number of downregulated genes were shown in blue bars, whilst the number of upregulated genes were shown in red bars.

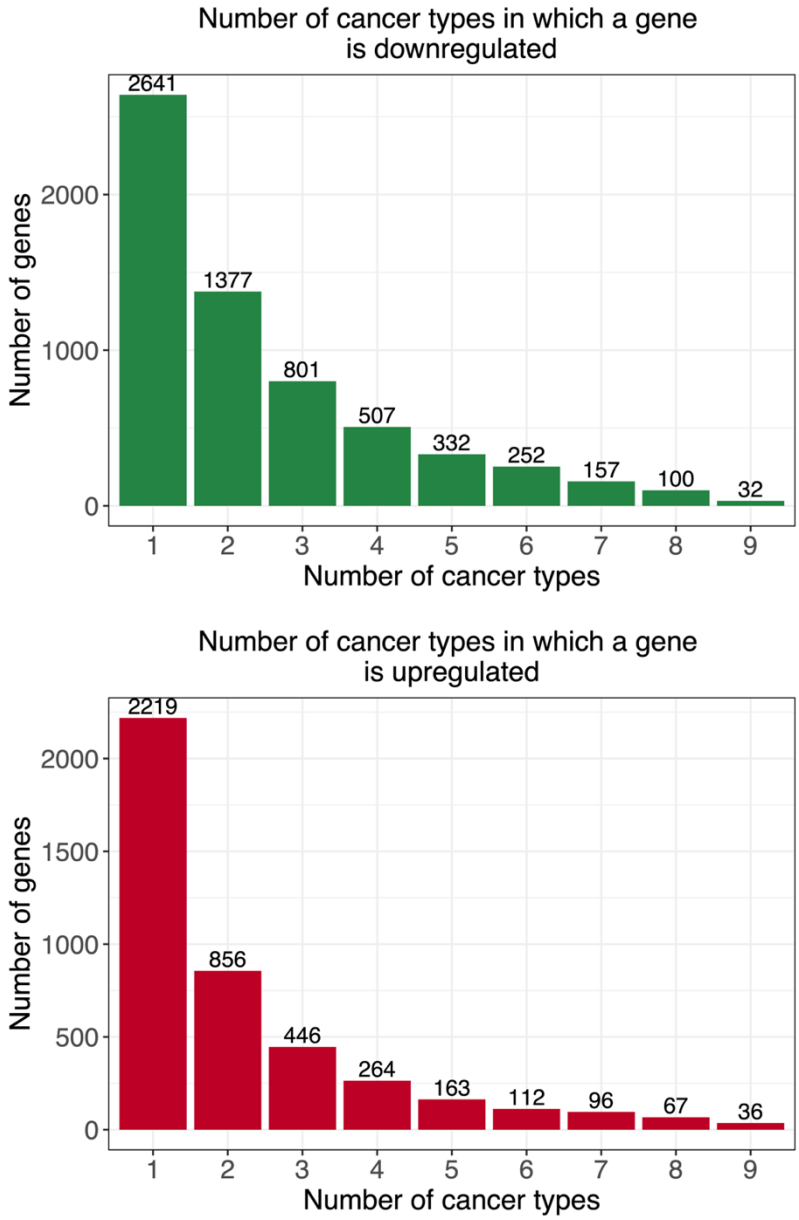


Figure 2.6. The number of tissues in which an age-DEG is downregulated or upregulated in.

(Top) The number of tissues in which a gene is significantly downregulated with age. **(Bottom)** The number of tissues in which a gene is significantly upregulated with age.

2.3.3 The relationship between differentially expressed genes with age and differentially expressed genes in cancer

After obtaining a list of cancer-DEGs, I examined the fold-change with age in non-cancerous tissues from GTEx of the cancer-DEGs. In other words, I investigated how cancer-DEGs change during normal ageing in a tissue-specific manner. In this analysis, I selected nine tissues with a matched TCGA project from the same tissue of origin, and these TCGA projects included more than ten matched normal tissue samples. The tissues I included in the analysis were the breast, colon, oesophagus, liver, lung, prostate, stomach, thyroid, and uterus.

To get an overview of transcriptomic changes during ageing and cancer in these tissues, I studied age-related changes in gene expression in GTEx normal samples of upregulated and downregulated genes in cancer. I observed a significant (adj. *p*-value < 0.05; two-sided Wilcoxon rank-sum test) higher fold-change with age in downregulated cancer-DEGs when compared with upregulated cancer-DEGs for most cancer types, with the opposite being observed in two tissues, THCA-thyroid and UCEC-uterus (**Figure 2.7**). This result indicates the tissue-specific pattern of gene expression changes in ageing and cancer. In general, gene expression changes in ageing and cancer were in the opposite direction, except for the thyroid and uterus.

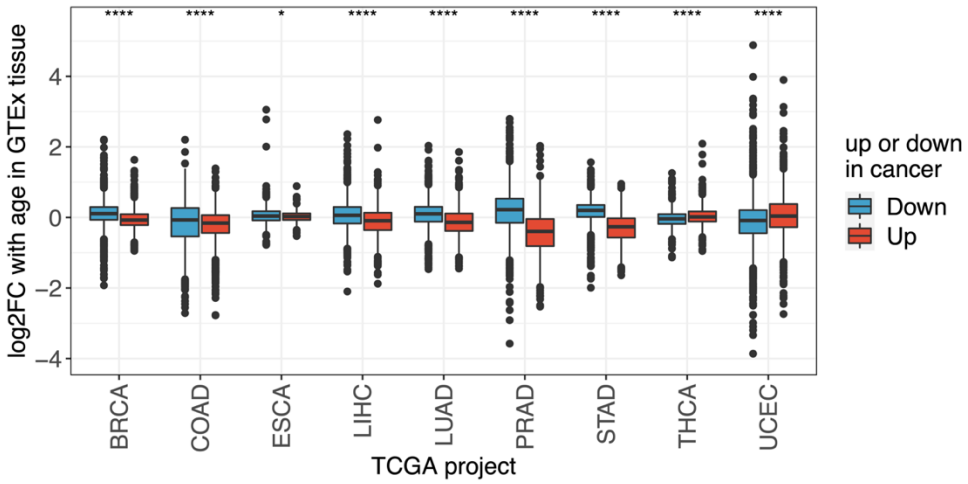


Figure 2.7. Fold changes with age in GTEx data of cancer-DEGs.

For each TCGA cancer type, downregulated cancer-DEGs and upregulated cancer-DEGs were examined for their fold changes with age in GTEx non-cancerous tissues. Differences of these age-related fold changes between downregulated cancer-DEGs and upregulated cancer-DEGs were tested using the two-sided Wilcoxon rank-sum test (*, adj. *p*-value < 0.05.; ****, adj. *p*-value < 0.0001).

I next overlapped these gene sets for each tissue to further compare age-DEGs and cancer-DEGs. For each tissue, gene overlap analyses were conducted in 4 cases:

1. Downregulated genes with age and in cancer (down with age - down in cancer).
2. Downregulated with age but upregulated genes in cancer (down with age - up in cancer).
3. Upregulated genes with age but downregulated genes in cancer (up with age - down in cancer).
4. Upregulated with age in cancer (up with age - up in cancer).

Consistent with the previous analysis, genes changed in the opposite direction between ageing and cancer significantly overlapped more often than genes changed in the same direction in breast, colon, oesophagus, lung, and prostate (Fisher's exact test, Benjamini-Hochberg correction) (**Figure 2.8**). The overlap was significant for genes changing in the same direction for the thyroid. There was no significant overlap in liver and stomach, which might be explained by the low number of age-DEGs in these tissues (16 genes in liver and 26 genes in the stomach, **Figure 2.2**). In the uterus, however, the overlap was significant in all cases.

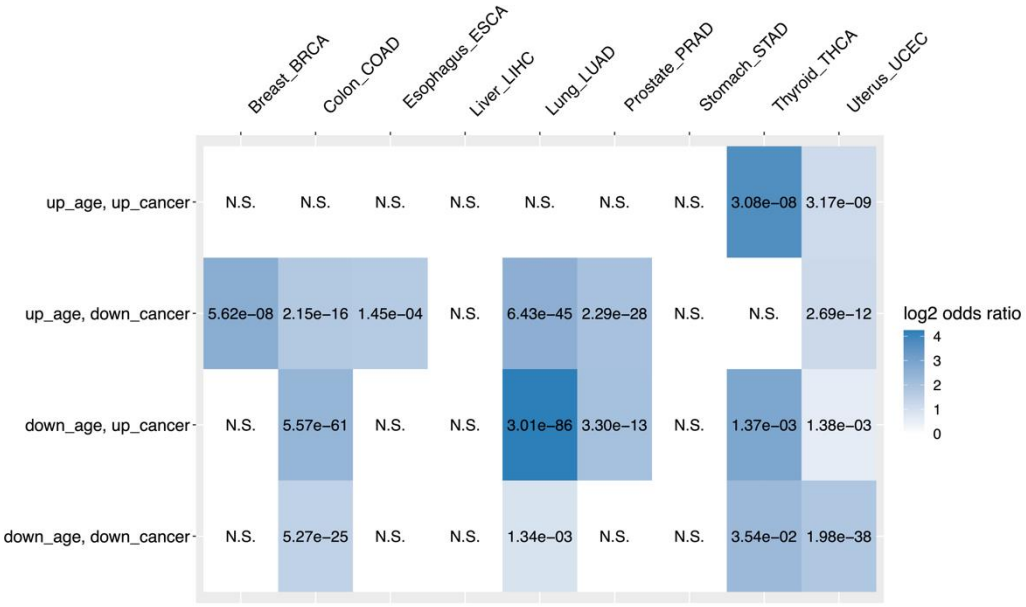


Figure 2.8. Tissue-specific overlap between age-DEGs and cancer-DEGs.

Overlap analyses between age-DEGs and cancer-DEGs were performed in a tissue-specific manner, using Fisher's exact test. The significant overlap gene sets were shown with their adj. *p*-values (adj. *p*-value < 0.05). N.S. denotes non-significant overlap. The colour shade represents the log2 odds ratio of the overlap.

I performed gene ontology enrichment analysis to better understand the biological processes related to overlapped differentially expressed genes in ageing and cancer. Six out of 20 significantly overlapping sets were significantly enriched in at least one GO term (**Figure 2.9**). Genes downregulated with age – upregulated in cancer, in colon and lung, were related to the cell cycle. Indeed, uncontrolled cell proliferation is one of the major hallmarks of cancer (Hanahan & Weinberg, 2011). On the other hand, cell cycle rates decline with age in human tissues, including colon (Tomasetti *et al.*, 2019). Interestingly, cell cycle-related terms were also enriched in genes upregulated with age – upregulated in cancer in the uterus. Uncontrolled cell proliferation in the ageing uterus often leads to endothelial hyperplasia and could lead to endometrial cancer (Damle *et al.*, 2013). Immune system-related terms were enriched in genes downregulated with age – downregulated in cancer in the colon. Because the immune system plays an important role in preventing cancer through immunosurveillance (Ribatti, 2017), compromised immune function with age could provide an immunosuppressive microenvironment, allowing cancer cells to evade immunosurveillance and thrive (Fulop *et al.*, 2010).

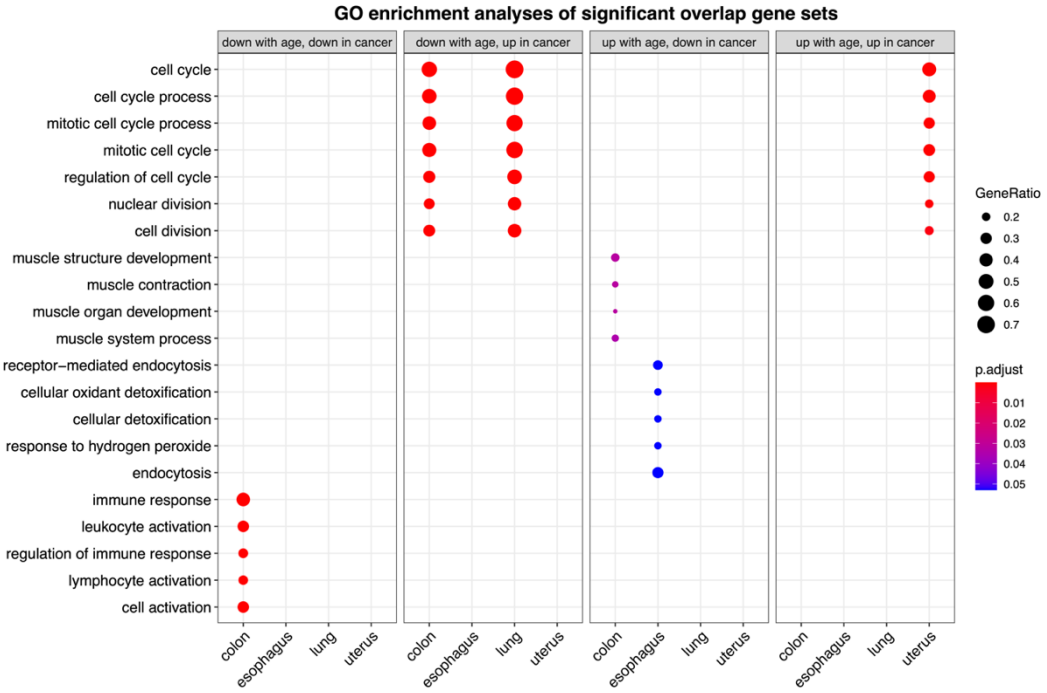


Figure 2.9. GO enrichment analysis of significantly overlapping gene sets between age-DEGs and cancer-DEGs.

The dot plot shows examples of significantly enriched terms from the overlapping gene sets (adj. *p*-value < 0.1). Colour represents adj. *p*-value. Dot size denotes gene ratio, the ratio of input genes annotated in a term.

Because the GTEx and TCGA data I obtained and used were processed differently, I ensured that our results were robust by reanalysing the relationship between age-DEGs and cancer-DEGs using recount2 data (Collado-Torres *et al.*, 2017). In recount2, GTEx (version 6) and TCGA data were pre-processed using the same pipeline, making them more comparable. By performing similar analyses, I confirmed that age-DEGs and cancer-DEGs change in the same direction in the thyroid and uterus while changing toward the opposite direction in the other tissues, confirming that our analyses were not confounded by the different pre-processing steps from GTEx and TCGA.

2.3.4 Cellular senescence signatures

I next performed a meta-analysis of 20 replicative senescence microarray datasets from the Gene Expression Omnibus (GEO) to identify cellular senescence signature genes (i.e., genes that are consistently overexpressed or underexpressed in senescence cells across several datasets). Datasets used for this analysis are shown in **Table 2.3**. There were 1,259 senescence signature genes, with 526 overexpressed and 734 underexpressed. I found that 442 of 1,259 senescence signature genes were significantly overlapped with the signature of replicative senescence provided by Hernandez-Segura et al. (Hernandez-Segura *et al.*, 2017) (p-value = 2.9×10^{161}; Fisher's exact test), indicating the reliability of our data. GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses revealed that underexpressed signatures were related to cell cycle and DNA repair. In contrast, the overexpressed signatures were linked to immune processes and the p53 signalling pathway, a well-established senescence pathway (**Figure 2.10**). This result suggests that I successfully obtained consistently up and downregulated genes in senescence cells. These genes may play a central role in the transcriptional programme of cellular senescence.

Table 2.3: GEO datasets for meta-analysis of cellular senescence signatures

GEO	Platform ID	References	Tissue	Cell Type/ Cell Line	Proliferating Cells	Senescent Cells
GSE17077	GPL1352	(Gruber *et al.*, 2010)	Annulus	-	8	8
GSE13330	GPL570	(Pazolli *et al.*, 2009)	Foreskin	BJ Fibroblasts	6	6
GSE49860	GPL5639	(Imai *et al.*, 2014)	-	Diploid Fibroblasts	1	1
GSE41714	GPL10558	(Kim *et al.*, 2013b)	Dermal	Diploid Fibroblasts	2 (2 days)	2 (30 and >30 days)
GSE37091	GPL6480	(Jong *et al.*, 2013)	Umbilical vein	Endothelial cell (HUVECS)	2	2
GSE54095	GPL6244	(Guerrero *et al.*, 2015)	Coronary Artery	Endothelial Cells	4	4
GSE3460	GPL2937	(Schwarze *et al.*, 2002)	Prostate	Epithelial Cells	2	2
GSE3460	GPL2947	(Schwarze *et al.*, 2002)	Prostate	Epithelial Cells	1	1
GSE3731	GPL2990	(Zhang *et al.*, 2003)	Mammary	Epithelial Cells (HMEC) (48R and 184)	8	8
GSE36640	GPL570	(Shah *et al.*, 2013)	-	Fibroblast/IMR90	5	5
GSE19018	GPL570	-	-	Fibroblast/IMR90	3	3
GSE28863	GPL5175	(Cao *et al.*, 2011)	-	Fibroblasts	12	12
GSE3730	GPL2990	(Zhang *et al.*, 2003)	-	Fibroblasts (WS1, WI38 and BJ)	13	13
GSE56530	GPL570	(Medeiros Tavares Marques *et al.*, 2017)	Umbilical Cord Vein	Mesenchymal Stem Cells	9	9
GSE35957	GPL570	(Benisch *et al.*, 2012)	Bone Marrow	Mesenchymal Stem Cells	5	5
GSE46019	GPL6244	(Frobel *et al.*, 2014)	Bone Marrow	Mesenchymal Stromal Cells	3	3
GSE15919	GPL1528	(Binet *et al.*, 2009)	-	MRC5 Fribroblasts	6	6
GSE10570	GPL4133	(Bhatia *et al.*, 2008)	Prostate	NHP8	3	3

GSE6762	GPL4693	(Johung *et al.*, 2007)	-	Primary Fibroblasts	4	4
GSE687	GPL506	(Hardy *et al.*, 2005)	Mammary Stroma	Primary Fibroblasts	12	12
GSE34303	GPL4133	(Ren *et al.*, 2013)	Bone Marrow	Stromal Cells	20	20

I constructed a protein-protein interaction (PPI) network based on over- and underexpressed senescence signature genes and their first neighbours to further characterise our cellular senescence signatures. The network consisted of 5,067 nodes (proteins) and 13,167 edges (interactions) (**Figure 2.11**). I analysed the network topology focusing on parameters including Degree, Betweenness Centrality (BC) and Closeness Centrality (CC). The degree denotes the number of interactions of a particular node (protein). BC is a way to represent the importance of each node on the flow of information in a network. BC is calculated by the unweighted shortest paths between all node pairs in the network that cross a particular node of interest. CC measures the average inverse distance of each node to all other nodes. Thus, CC represents the centrality of each node in the network. The top 10 proteins with the highest Degree, BC, and CC are shown in **Table 2.4**.

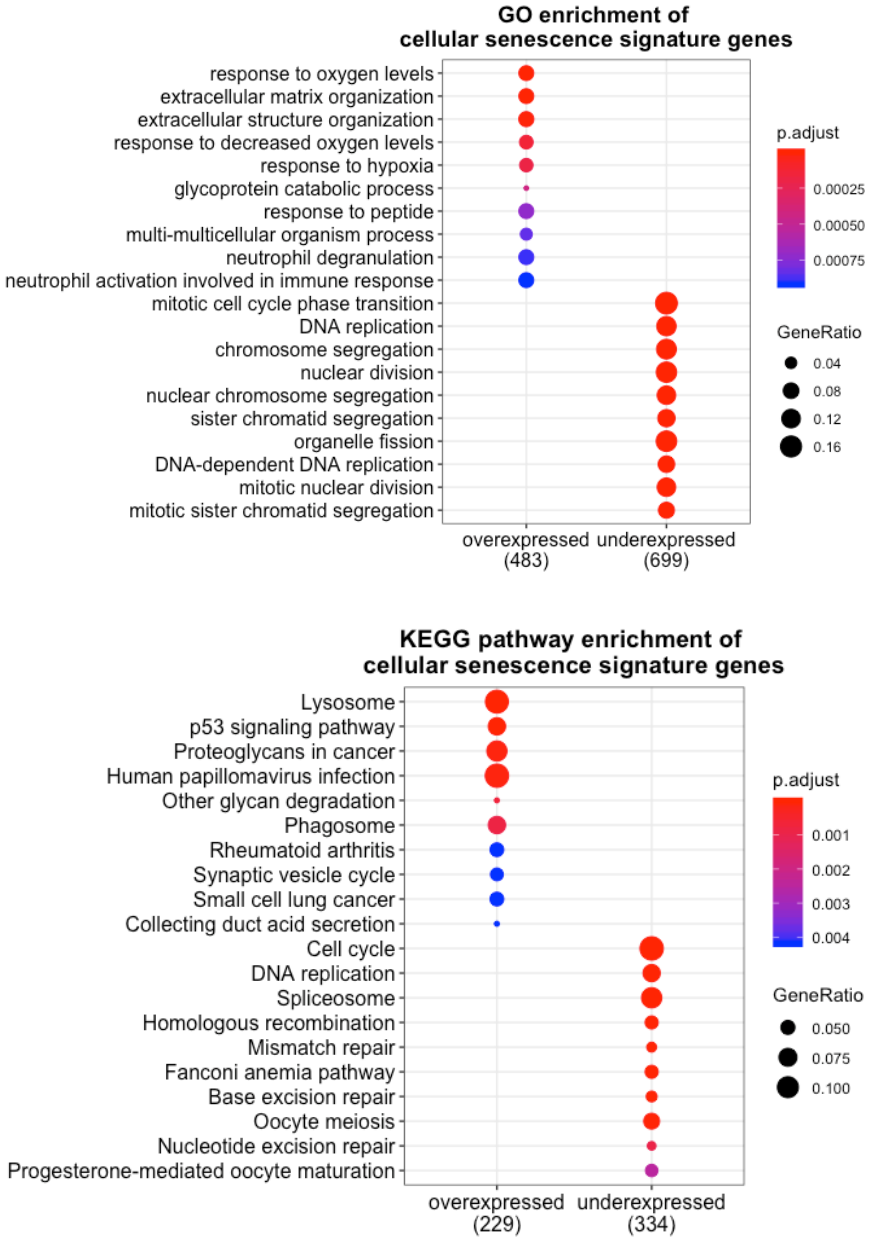


Figure 2.10. GO and KEGG pathway enrichment analysis of cellular senescence signatures.

(Top) GO and **(Bottom)** KEGG pathway enrichment analysis of overexpressed and underexpressed cellular senescence signatures. The dot plot shows examples of significantly enriched terms (adj. *p*-value < 0.05). Colour represents adj. *p*-value. Dot size denotes gene ratio.

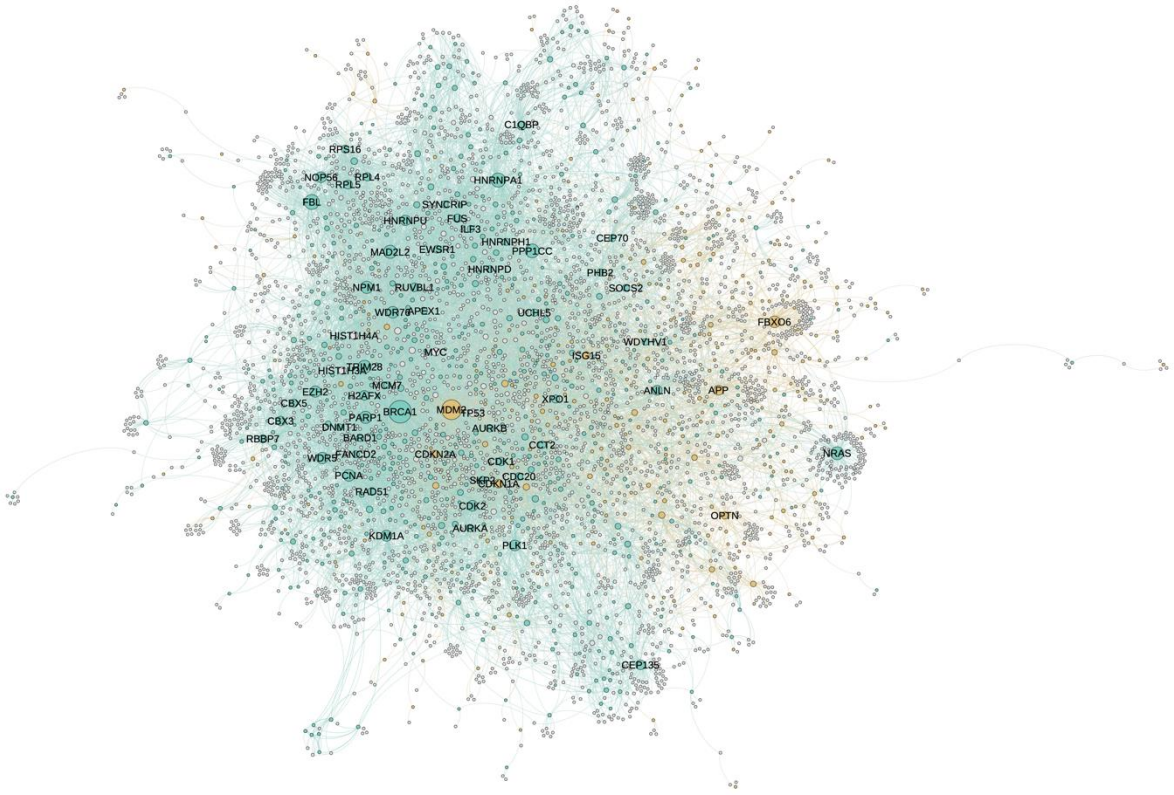


Figure 2.11. Protein-protein interaction (PPI) network of cellular senescence signature genes.

The PPI network was constructed from known interactions from the BioGrid database of meta-analysis-derived cellular senescence signature genes and their first-order neighbours. The network was visualised using Gephi software. Node size corresponds to a degree. Node colour represents senescence signature type: light brown – overexpressed signature, light blue – underexpressed signature, and grey – non-senescence signature.

Three proteins (BRCA1, MDM2 and TRIM28) were shared between the top 10 Degree, BC, and CC, indicating the critical role in the senescence signature network of these three proteins. BRCA1 was the protein with the highest degree (273 connections), BC (0.097), and CC (0.37), suggesting that this protein is a hub of the network. BRCA1 acts as a tumour suppressor gene. It plays an important role in maintaining genomic stability by involving in the DNA repair process. BRCA1 deficiency resulted in increased senescence (Santarosa *et al.*, 2009). In other words, BRCA1 is a senescence inhibitor. Thus, our result that BRCA1 is an underexpressed senescence signature corroborates this previous finding, suggesting that BRCA1 prevents cells from undergoing senescence and the downregulation of this gene is necessary for cells to turn into senescence cells. MDM2 was the second highest in all three parameters. MDM2 is a proto-

oncogene that inhibits the TP53 tumour suppressor gene. A recent study reported that nutlin-3a and MI-63, small molecule inhibitors of MDM2, attenuated the SASPs and thus limited the cancer-promoting effect of senescent cells through SASPs. (Wiley *et al.*, 2018). TRIM28, an underexpressed senescence gene on our list, was also among the top 10 in all three parameters. Interestingly, TRIM28 was reported as an inducer of oncogene-induced senescence (OIS). TRIM28 depletion partially prevented cell arrest during OIS (Santos & Gil, 2014). Since our meta-analysis was performed on replicative senescence datasets, the expression and role of TRIM28 may likely be different between replicative senescence and OIS.

TP53, MYC, and ESR1 were not in our list of senescence signatures; however, they were among the top 10 CC proteins located in the centre of the network. In fact, both TP53 and MYC are in the CellAge database of cellular senescence genes (Avelar *et al.*, 2020). Therefore, it is likely that these genes, together with other genes in CellAge, could coordinate the gene expression of cellular senescence. TP53 is a well-known tumour suppressor gene and the most frequently mutated gene in cancer (Martinez-Jimenez *et al.*, 2020). Because of its central role in DNA damage response, TP53 is often referred to as "the guardian of the genome". Stresses such as DNA damage, telomere shortening, and the oncogenic signal can activate p53 signalling through ATM kinase. The p53 will then trigger p21 and E2F7, suppressing the cell cycle, leading to cell cycle arrest (Glass *et al.*, 2013). Next, decreased MYC signalling could induce cellular senescence mediated by the cyclin-dependent kinase inhibitor $p16^{INK4a}$ (Guney *et al.*, 2006). ESR1 is not on cellular senescence signature genes and CellAge senescence-related genes, but it is also among the top 10 highest closeness centrality genes in the senescence PPI network. Thus, it would be interesting to study this gene in the context of cellular senescence, in addition to its known role in ageing (Tacutu *et al.*, 2018).

Taken together, I constructed a PPI network from the meta-analysis derived cellular senescence signature genes and thus identified interconnected genes relating to cellular senescence transcriptional programmes.

Table 2.4: Top 10 Degree, BC, and CC genes in PPI network of cellular senescence signature genes.

Degree	Betweenness Centrality	Closeness Centrality
BRCA1 (Under)	BRCA1 (Under)	BRCA1 (Under)
MDM2 (Over)	MDM2 (Over)	MDM2 (Over)
FBL (Under)	NRAS (Under)	TRIM28 (Under)
HNRNPA1 (Under)	FBXO6 (Under)	TP53 (None)
PPP1CC (Under)	PPP1CC (Under)	NPM1 (Under)
MAD2L2 (Under)	MAD2L2 (Under)	MYC (None)
TRIM28 (Under)	TRIM28 (Under)	PARP1 (Under)
NRAS (Under)	KDM1A (Under)	ESR1 (None)
NPM1 (Under)	APP (Over)	PLK1 (Under)
PARP1 (Under)	HNRNPA1 (Under)	PCNA (Under)

2.3.5 The relationship between cellular senescence signatures, differentially expressed genes with age, and differentially expressed genes in cancer

To investigate how senescence signature genes change during ageing, I evaluated fold changes of overexpressed and underexpressed signatures across 50 years of age from our GTEx linear regression analysis (**Figure 2.12**). I found that overexpressed senescence signatures had a significant higher fold change with age than underexpressed senescence signatures in 17 out of 26 tissues (adjusted p-value < 0.05, two-sided Wilcoxon rank-sum test). Only four tissues, including the adrenal gland, nerve, salivary gland, and uterus, showed an opposite pattern that underexpressed signatures had a significant higher fold change with age than overexpressed signatures. Therefore, changes in gene expression during ageing coincide with cellular senescence gene expression programmes in most tissues. These results might reflect an accumulation of senescence cells in most ageing tissues.

I next performed a similar analysis to investigate the relationship between cellular senescence signatures and cancer gene expression changes. Not surprisingly, I found that overexpressed senescence genes were strongly downregulated in cancer, while underexpressed senescence genes were greatly upregulated in cancer in most cancer types (**Figure 2.13**). This observation is consistent with the canonical anti-cancer role of cellular senescence. Only in thyroid cancer

that there was no significance different between the fold change in tumours of overexpressed senescence signatures and changes in underexpressed senescence signatures.

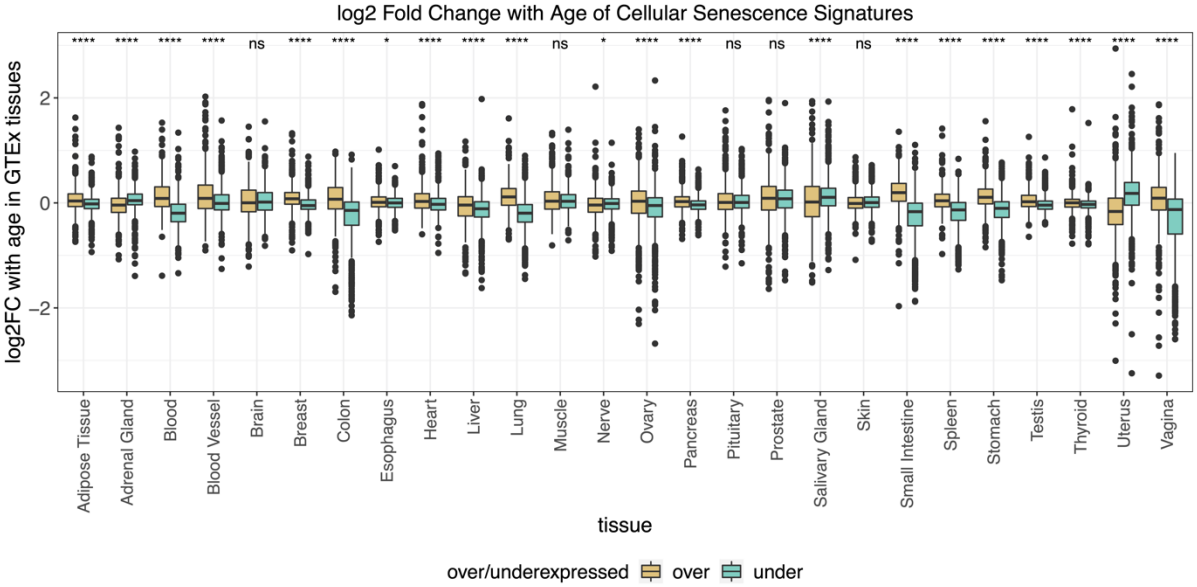


Figure 2.12. Fold changes with age of cellular senescence signatures.

Overexpressed and underexpressed senescence signatures were examined for their fold changes with age in 26 GTEx tissues. Differences of these age-related fold changes between overexpressed and underexpressed senescence signature genes were tested using the two-sided Wilcoxon rank-sum test (*, adj. *p*-value < 0.05.; ****, adj. *p*-value < 0.0001; ns, not significant).

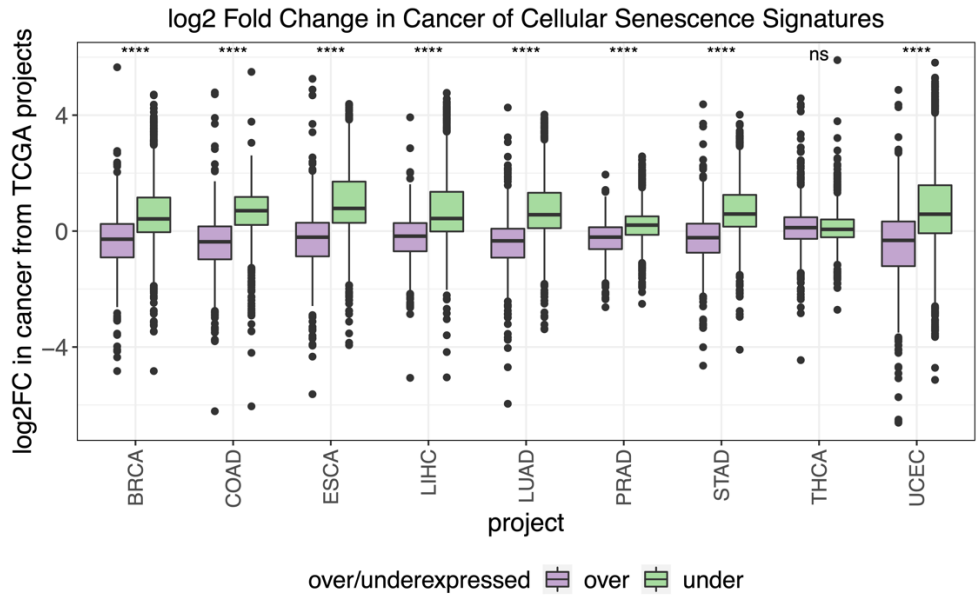


Figure 2.13. Fold changes in cancer of cellular senescence signatures.

Overexpressed and underexpressed senescence signatures were examined for fold changes between cancer and matched normal tissues in 9 TCGA cancer types. Differences of these cancer-related fold changes between overexpressed and underexpressed senescence signature genes were tested using the two-sided Wilcoxon rank-sum test (****, adj. *p*-value < 0.0001; ns, not significant).

I further confirmed these findings using the overlap analysis in nine tissues used for ageing and cancer comparison. Underexpressed senescence signatures significantly overlapped with genes downregulated age-DEGs in the colon and lung, while overexpressed senescence signatures overlapped with upregulated age-DEGs in colon, lung, prostate, and thyroid (Fisher's exact test, **Figure 2.14A-B**). Only the uterus displayed opposite results, in line with the up-regulation in the ageing uterus of cell cycle-related genes (**Figure 2.9**). For cancer genes, an opposite solid trend between cancer-DEGs and senescence signatures emerged (**Figure 2.14C-D**), consistent with the anti-cancer role of senescence. Only in thyroid cancer (THCA) did the upregulated genes in cancer significantly overlap with overexpressed senescence genes. These overlapped genes are related to cell cycle regulation, cell adhesion, and apoptosis.

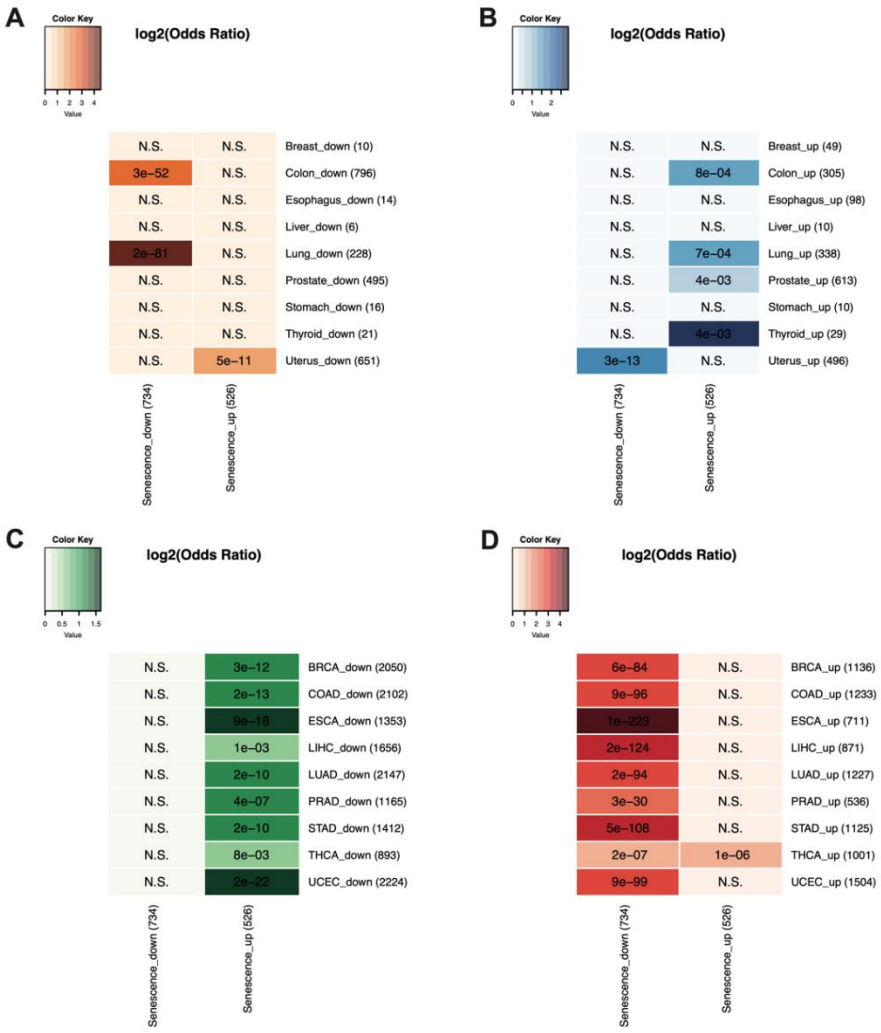


Figure 2.14. Overlap between cellular senescence signature genes, age-DEGs, and cancer-DEGs.

Fisher's exact test was used to investigate the overlap between cellular senescence signatures and **(A)** downregulated age-DEGs, **(B)** upregulated age-DEGs, **(C)** downregulated cancer-DEGs, and **(D)** upregulated cancer-DEGs. The significant overlap gene sets were shown with their adj. *p*-values (adj. *p*-value < 0.05). N.S. denotes non-significant overlap. The colour shade represents the log2 odds ratio of the overlap.

2.4 Discussion

Because age is the most important risk factor for cancer, it is imperative to better understand how ageing processes relate to cancer biology. This chapter employed large-scale gene expression datasets from several sources to identify tissue-specific age-related gene expression changes and asked how these changes associate with transcriptomic alterations in cancer. The overlap gene sets between differentially expressed genes with age in non-cancerous tissues and differentially expressed genes between tumours and normal were enriched in numerous biological processes, most importantly in cell cycle regulation. I further identified cellular senescence signature genes, genes that consistently changed in cellular senescence across several datasets and demonstrated that they change in concert with ageing genes and are strongly opposed to cancer genes.

2.4.1 Tissue-specific differentially expressed genes during ageing

The advent of assays to measure global transcriptome, initially microarray and later RNA-seq, has opened an unprecedented opportunity to study the basis of gene expression changes in numerous biological processes, including ageing. Over the past two decades, gene expression analysis has previously been employed to study changes associated with human ageing in many tissues such as blood, brain, skin, muscle, and liver. However, many of these studies contained only a small sample size. The large-scale consortium-based data such as GTEx, thus, has provided a unique opportunity to conduct a gene expression analysis in relation to ageing with a large sample size across a wide range of tissues. A previous study has used GTEx data version 3 to study age-related gene expression in 9 tissues. They found the tissue-specific pattern of ageing gene expression. In several cases, they also reported the overlap between ageing genes and disease genes, most prominently for tissues affected by the corresponding diseases (Yang *et al.*, 2015).

Here, using a more recent version of GTEx (version 7), I am able to identify tissue-specific age-related gene expression changes in 26 tissues with more than 90 samples. Consistent with previous reports, age-related differentially expressed genes are essentially tissue specific as most of the significant upregulated or downregulated genes were found only in one tissue. Biological processes related to changes in gene expression with ageing in several tissues are linked to known ageing-related mechanisms, including loss of cell cycle, a decline in developmental pathways, decreased mitochondria function and increased inflammation. Age-

related cell cycle decline has recently been shown in human colonic, duodenal, oesophageal, and posterior ethmoid sinonasal tissues (Tomasetti *et al.*, 2019). I also found a reduced expression of genes related to tissue-specific functions, such as decreased neuronal function in the ageing brain. Together, I identified human tissue-specific gene expression changes in biological processes in relation to age.

While our results have shed light on the gene expression pattern changes in human ageing across several tissues, several important questions remain to be elucidated. First, these results were obtained from the bulk analysis. Thus, I cannot distinguish whether changes in gene expression result from changes in cellular population in the tissue with age or the intrinsic differences in transcriptome within a cell type. Indeed, a rapidly growing field of single-cell genomics has started to shed light on this issue in model organisms and humans. For example, Enge *et al*. performed scRNA-seq and reported increased levels of transcriptional noise and potential fate drift of islet endocrine cells in aged human pancreas (Enge *et al.*, 2017). Recently, a multi-tissue scRNA-seq has been conducted in mice by the *Tabula Muris* Consortium, creating a reference atlas of the murine single-cell ageing transcriptome (Tabula Muris, 2020). This study denoted several changes in relation to age, such as increased mutation load and changes in the immune system. It is expected that a similar atlas in humans should be released in the next coming years. However, due to the expensive nature of single-cell techniques, it is difficult or almost not possible at the moment to create large-scale scRNA-seq data covering hundreds of donors.

Given a large number of studies and the consistency between them in terms of biological processes associated with ageing gene expression changes, the current important questions are moving away from ‘what’ genes in which their expression changes during ageing to ‘how’ and ‘why’ the changes occur. Epigenetic alteration is one of the hallmarks of ageing (Lopez-Otin *et al.*, 2013). Gene expression can be regulated by numerous epigenetic processes such as DNA methylation and histone modifications. These epigenetic mechanisms have extensively been studied in the context of ageing (Pal & Tyler, 2016). However, epigenetic changes and gene expression changes are not always consistent. Several studies reported that age-associated DNA methylation changes in humans could explain only a small fraction of gene expression changes (Reynolds *et al.*, 2014; Steegenga *et al.*, 2014; Marttila *et al.*, 2015). Furthermore, a recent work performed a multi-omics analysis of the ageing bone marrow mesenchymal stromal cells. It showed a poor correlation between age-associated changes in gene expression,

chromatin accessibility, and protein expression (Lai *et al.*, 2021). Therefore, the major challenge in the coming years is pinpoint mechanisms underlying changes in age-related gene expression.

2.4.2 A human tissue-specific relationship between gene expression changes in ageing and cancer

One of the major findings in this chapter is the tissue-specific pattern of the relationship between ageing and cancer transcriptome. More specifically, I found that human ageing and cancer gene expression change toward the opposite direction in most tissues, except the thyroid and uterus. The opposite changes between ageing and cancer gene expression are in line with a recent study reporting that ageing transcriptome shifts away from cancer and moves toward degenerative diseases (Aramillo Irizar *et al.*, 2018). Our results strengthen this observation. In addition, I underscore strong exceptions in the thyroid and uterus, which could reflect an intrinsic ageing transcriptional programme of the tissue.

One possible interpretation of this opposite ageing and cancer transcriptome is that molecular changes during ageing processes may oppose cancer development. Only cells with oncogenic mutations and evade ageing phenotypes could thrive and develop into clones and eventually tumours. For example, cells carrying mutations in a component of cellular senescence pathways such as TP53 could avoid cell cycle arrest and develop into cancer. Another potential interpretation for our results, from an evolutionary perspective, is that ageing-related changes in the tissue microenvironment, leading to the decrease in tissue robustness, might provide a selective advantage for cells harbouring oncogenic mutations. Supporting this hypothesis, a previous study showed that ageing-associated functional decline of aged B-lymphoid progenitors permitted a tremendous competitive advantage for B progenitors carrying Bcr-Abl translocations, facilitating the progression of Bcr-Abl-driven leukaemias (Henry *et al.*, 2010). Another study also supported this idea by showing that aged mice transplant with haematopoietic stem cells expressing AML1-ETO fusion protein, the most frequent genetic alteration in acute myeloid leukaemia, contained a higher frequency of progenitor cells expressing AML1-ETO than younger mice, suggesting that ageing allow a pervasive environment for cancer cells to thrive (Vas *et al.*, 2012). Accumulating evidence highlights the crucial roles of ageing-related microenvironment changes in cancer progression (Fane & Weeraratna, 2020). Furthermore, tumourigenesis can be a tissue-specific process (Schneider *et*

al., 2017). In the future, it is necessary to extend these experimental works to distinct cancer types to better understand the influence of ageing-related changes in cancer in a tissue-specific context.

2.4.3 Cellular senescence signature genes and their relationship with ageing and cancer transcriptome

Cellular senescence is a state of irreversible cell cycle arrest in response to stress and damage, preventing the cell from transforming into cancer. Conversely, senescent cells secrete various molecules, such as inflammatory cytokines, chemokines, and growth factors, collectively termed senescence-associated secretory phenotypes (SASP), which could help promote tumour progression (Campisi, 2013). Accumulation of senescent cells with age, therefore, could contribute to cancer progression. To better understand the relationship between cellular senescence, ageing, and cancer, I derived replicative senescence signature genes using a meta-analysis of 20 microarray datasets. These datasets consist of different cell types; thus, our signature should represent cellular senescence transcriptional programme irrespective of cell types. Our senescence signatures are consistent but do not entirely overlap, with signature for replicative senescence from a recent study (Hernandez-Segura *et al.*, 2017). This is due to differences in terms of datasets used in the analysis and Hernandez-Segura's signature for replicative senescence used only fibroblasts in their analysis. Nevertheless, our gene list compliments this study and will serve as a resource for the research community.

Senescent cells accumulate with age in mice (Yousefzadeh *et al.*, 2020) and humans (Tuttle *et al.*, 2021). I showed that upregulated and downregulated age-related gene expression is associated with overexpressed and underexpressed cellular senescence signatures, respectively. This is consistent with an age-related accumulation of senescent cells. The degree of increased overexpressed senescence signatures and decreased underexpressed senescence signatures vary from tissue to tissue. For example, very strong overlapped between ageing-related genes and cellular senescence signatures were found in the colon and lung. This result could be due to different possibilities. First, senescent cells accumulate at different rates between tissues. Second, several cell types within the tissue exhibit gene expression change toward senescence, but with various gene expression changes in different cell types in different tissues. Indeed, scRNA-seq data in ageing human tissues are needed to provide a clearer picture on this issue. A good example of this is the recent scRNA-seq study in mice showing an

increase in cell population expressing *Cdkn2a*, which code for protein $p16^{INK4a}$, in aged mice. Traditionally, senescent cells are defined by a few markers such as $p16^{INK4a}$, $p21^{CIP1}$, and p53. However, a consensus marker for cellular senescence is currently lacking (Gorgoulis *et al.*, 2019). In the future, it would be interesting to study senescent cells using gene sets instead of a gene as markers, such as cellular senescence signatures.

Finally, our result reveals that cellular senescence signatures strongly oppose cancer gene expression. This result is not surprising, given that cancer cells proliferate uncontrollably while senescent cells exhibit growth and cell cycle arrest. Only in thyroid cancer, I find a significant overlap between upregulated cancer genes and overexpressed cellular senescence signatures. Interestingly, a recent study suggests that senescent tumour cells presented in the front region of collective invasion and are involved in thyroid cancer metastasis (Kim *et al.*, 2017). Further studies are required to fully understand the role of cellular senescence and senescence-related genes in thyroid cancer.

Overall, I have explored the relationship between ageing, cancer, and cellular senescence through the transcriptomic lens in this chapter. I showed that ageing and cancer gene expression patterns changed in the opposite direction (**Figure 2.15**). Moreover, cellular senescence genes are altered in the same way as ageing in human tissues and the opposite to cancer transcriptome (**Figure 2.15**). These results provide novel insights into the complex relationship between ageing, cancer, and cellular senescence.

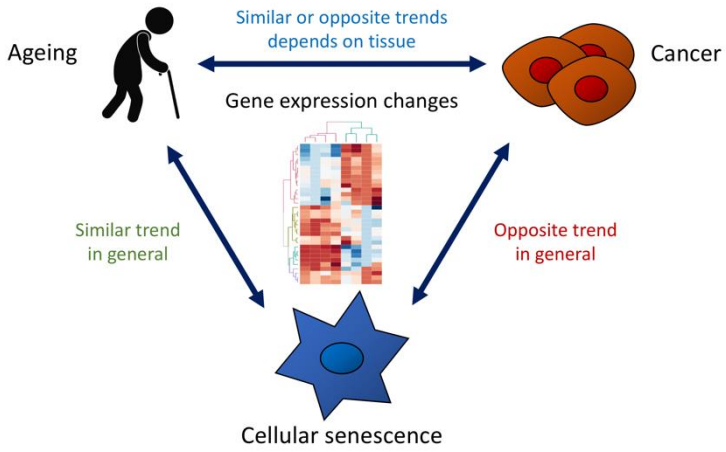


Figure 2.15. The gene expression relationship between ageing, cancer, and cellular senescence.

Chapter 3

The age-associated multi-omic landscape across cancers

3.1 Introduction

Age is the biggest risk factor for cancer, as cancer incidence and mortality rates increase exponentially with age in most cancer types (de Magalhaes, 2013). As the ageing population continues to rise, a better understanding of the relationship between ageing and cancer is critically needed. In addition, the disparities between cancer from young and aged patients have also been observed. (Van Herck *et al.*, 2021). For instance, breast cancer in younger patients tends to be more aggressive and associated with poorer survival (Anders *et al.*, 2008), while the prognosis is worse in older ovarian cancer patients (Maas *et al.*, 2005). However, how molecular alteration in tumours differs among patients of different ages has not been systematically characterised. Cancer arises through the interplay between somatic mutations and selection in a Darwinian-like process (Nowell, 1976; Laconi *et al.*, 2020). Thus, apart from the accumulation of mutations with age (Tomasetti *et al.*, 2013; Alexandrov *et al.*, 2015; Milholland *et al.*, 2015), microenvironment changes during ageing could also play a role in carcinogenesis (Chatsirisupachai *et al.*, 2019; Fane & Weeraratna, 2020; Laconi *et al.*, 2020). I, therefore, hypothesise that, due to the differences in selective pressures from tissue microenvironmental changes with age, tumours arising from patients across different ages might harbour distinct molecular landscapes. Consequently, some molecular changes might present more or less common as a function of age.

Recently, several studies have investigated molecular differences in the cancer genome in relation to clinical factors, including gender (Yuan *et al.*, 2016; Li *et al.*, 2018) and race (Yuan *et al.*, 2018; Sinha *et al.*, 2020). These studies demonstrated gender- and race-specific biomarkers, actionable target genes and provided clues to understanding the biology behind the disparities in cancer incidence, aggressiveness and treatment outcome across patients from different backgrounds. Although differences in genomic alterations between childhood and adult cancers have been systematically characterised (Grobner *et al.*, 2018; Ma *et al.*, 2018), the age-related genomic landscape across adult cancers remains elusive. Specific age-associated molecular landscapes have been reported in the cancer genome of several cancer

types. For example, older breast cancer patients are associated with higher *CDH1* and *PIK3CA* mutations, while younger patients are diagnosed with higher frequencies of *TP53* and *GATA3* mutations (Azim *et al.*, 2015; Liao *et al.*, 2015; Mealey *et al.*, 2020). In colorectal cancer, somatic mutations in *TP53* and *CTNNB1* are more common in younger patients, while mutations in *APC*, *KRAS*, and *BRAF V600* are found more often in older patients (Berg *et al.*, 2010; Lieu *et al.*, 2019; Willauer *et al.*, 2019). Although these studies are informative, they focused mainly on a single cancer type and some molecular data types, leaving a gap for the pan-cancer multi-omic survey of the association between age and cancer molecular data.

In this chapter, I took advantage of the pan-cancer multi-omic datasets from TCGA to systematically investigate age-related differences in genomic instability, SCNAs, somatic mutations, pathway alterations, gene expression, and DNA methylation. I show that, in general, genomic instability and mutations frequency increase with age. I identify several age-associated genomic alterations in cancers, particularly low-grade glioma and endometrial carcinoma. Moreover, I also demonstrate that age-related DNA methylation changes partly control age-related gene expression changes and that these changes are linked to numerous biological processes. Works in this chapter have been published in *Nature Communications* (Chatsirisupachai *et al.*, 2021).

3.2 Materials and Methods

3.2.1 Data acquisition

Publicly available copy-number alteration seg files (nocnv_hg19.seg), normalised mRNA expression in RSEM (.rsem.genes.normalized_results TCGA files from the legacy archive, aligned to hg19), and clinical data (XML files) from TCGA were downloaded using *TCGAbiolinks* (version 2.14.1) (Colaprico *et al.*, 2016). The mutation annotation format (MAF) file was downloaded from the TCGA Mutation Calling in Multiple Cancers (MC3) project (Ellrott *et al.*, 2018) (https://gdc.cancer.gov/about-data/publications/mc3-2017). The somatic alterations in 10 canonical oncogenic pathways across TCGA samples were obtained from a previous study by Sanchez-Vega *et al.* (Sanchez-Vega *et al.*, 2018). The TCGA Illumina HumanMethylation450K array data (in β-values) was downloaded from Broad GDAC Firehose (http://gdac.broadinstitute.org/). The allele-specific copy number, tumour ploidy, tumour purity that were estimated using ASCAT (version 2.4.2) (Van Loo *et al.*, 2010) on hg19 SNP6 arrays with penalty=70 were obtained from previous studies (Alexandrov *et al.*, 2016; Martincorena *et al.*, 2017), available at (https://github.com/Crick-CancerGenomics/ascat/tree/master/ReleasedData/TCGA_SNP6_hg19). Only samples that have these profiles available were included in subsequent analyses. WGD duplication was determined using a fraction of the genome with LOH and ploidy information. Genomic instability (GI) scores have been computed as a fraction of genomic regions that are not in 1+1 (for non WGD tumours) or 2+2 (for WGD tumours) statuses. For each data type and each cancer type, the summary of the numbers of TCGA samples included in the analysis are presented in **Table 3.1**. The clinical variables included in the analyses are presented in **Table 3.2**.

Table 3.1: Summary of TCGA cancer type and number of samples used in each analysis

Cancer type	Abbreviation	GI, LOH, WGD	SCNAs	Mutations (hypermutated and MSI-H removed)	Pathway alterations	Gene expression	DNA methylation
Adrenocortical carcinoma	ACC	89	89	89 (88)	76	77	78
Bladder Urothelial Carcinoma	BLCA	370	369	369 (364)	361	366	370
Breast invasive carcinoma	BRCA	1015	1011	954 (946)	922	1011	719
Cervical squamous cell carcinoma and endocervical adenocarcinoma	CESC	287	287	271 (263)	264	284	287
Cholangiocarcinoma	CHOL	35	35	35 (35)	35	35	35
Colon adenocarcinoma	COAD	411	411	374 (303)	323	410	278
Lymphoid Neoplasm Diffuse Large B-cell Lymphoma	DLBC	42	42	32 (32)	32	42	42
Oesophageal carcinoma	ESCA	176	176	176 (174)	165	175	176
Glioblastoma multiforme	GBM	489	489	356 (354)	116	137	259
Head and Neck squamous cell carcinoma	HNSC	489	489	472 (469)	459	481	489
Kidney Chromophobe	KICH	66	66	66 (66)	65	66	66
Kidney renal clear cell carcinoma	KIRC	496	496	343 (343)	331	493	296
Kidney renal papillary cell carcinoma	KIRP	228	228	222 (222)	215	228	213
Acute Myeloid Leukaemia	LAML	126	121	55 (54)	101	102	121
Brain Lower Grade Glioma	LGG	488	488	484 (484)	482	488	488
Liver hepatocellular carcinoma	LIHC	355	355	342 (340)	334	349	355
Lung adenocarcinoma	LUAD	460	460	456 (438)	446	456	402
Lung squamous cell carcinoma	LUSC	460	460	444 (437)	426	457	336
Mesothelioma	MESO	82	82	77 (77)	77	82	82
Ovarian serous cystadenocarcinoma	OV	556	556	397 (395)	173	288	545
Pancreatic adenocarcinoma	PAAD	133	133	130 (129)	113	127	132
Pheochromocytoma and Paraganglioma	PCPG	165	157	165 (164)	154	165	165
Prostate adenocarcinoma	PRAD	434	434	434 (432)	425	434	434
Rectum adenocarcinoma	READ	152	152	132 (125)	109	151	95
Sarcoma	SARC	229	229	213 (211)	209	227	229
Skin Cutaneous Melanoma	SKCM	434	434	432 (340)	332	433	434
Stomach adenocarcinoma	STAD	388	388	385 (345)	340	365	341
Testicular Germ Cell Tumours	TGCT	129	129	124 (124)	123	129	129
Thyroid carcinoma	THCA	260	260	249 (248)	244	259	258
Thymoma	THYM	76	76	76 (76)	73	73	76
Uterine Corpus Endometrial Carcinoma	UCEC	434	434	421 (282)	406	432	360
Uterine Carcinosarcoma	UCS	52	52	52 (51)	52	52	52
Uveal Melanoma	UVM	72	72	72 (72)	72	72	72
Total		9678	9660	8899 (8448)	8055	8946	8414

Table 3.2: Clinical variables included in the analysis model for each cancer type (N = not included; Y = included)

Abbreviation	Age	Gender	Tumour purity	Race	Pathologic stage	Histologic grade	Subtype	Smoking history	Alcohol consumption	Hepatitis	Figo stage	ER statue	Gleason score
ACC	Y	Y	Y	N	Y	N	N	N	N	N	N	N	N
BLCA	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N
BRCA	Y	Y	Y	Y	Y	N	N	N	N	N	N	Y	N
CESC	Y	Y	Y	N	N	Y	N	N	N	N	Y	N	N
CHOL	Y	Y	Y	Y	Y	N	N	Y	N	N	N	N	N
COAD	Y	Y	Y	N	Y	N	Y	N	N	N	N	N	N
DLBC	Y	Y	Y	Y	N	N	Y	N	N	N	N	N	N
ESCA	Y	Y	Y	N	N	Y	N	N	Y	N	N	N	N
GBM	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
HNSC	Y	Y	Y	Y	N	Y	N	Y	Y	N	N	N	N
KICH	Y	Y	Y	Y	Y	N	N	Y	N	N	N	N	N
KIRC	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N
KIRP	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N
LAML	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
LGG	Y	Y	Y	Y	N	Y	N	N	N	N	N	N	N
LIHC	Y	Y	Y	Y	Y	Y	N	N	Y	Y	N	N	N
LUAD	Y	Y	Y	Y	Y	N	N	Y	N	N	N	N	N
LUSC	Y	Y	Y	N	Y	N	N	Y	N	N	N	N	N
MESO	Y	Y	Y	Y	Y	N	Y	N	N	N	N	N	N
OV	Y	Y	Y	Y	N	Y	N	N	N	N	Y	N	N
PAAD	Y	Y	Y	Y	Y	Y	N	N	Y	N	N	N	N
PCPG	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
PRAD	Y	Y	Y	Y	N	N	N	N	N	N	N	N	Y
READ	Y	Y	Y	N	Y	N	Y	N	N	N	N	N	N
SARC	Y	Y	Y	Y	N	N	Y	N	N	N	N	N	N
SKCM	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N
STAD	Y	Y	Y	N	Y	Y	N	N	N	N	N	N	N
TGCT	Y	Y	Y	Y	Y	N	Y	N	N	N	N	N	N
THCA	Y	Y	Y	N	Y	N	Y	N	N	N	N	N	N
THYM	Y	Y	Y	Y	N	N	N	N	N	N	N	N	N
UCEC	Y	Y	Y	Y	N	Y	N	N	N	N	Y	N	N
UCS	Y	Y	Y	N	N	N	N	N	N	N	Y	N	N
UVM	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	N

3.2.2 Statistical analysis and visualisation

Simple linear regression and multiple linear regression adjusting for clinical variables were performed using the *lm* function in R to access the relationship between age and continuous variables of interest. Simple logistic regression to investigate the association between age and binary response (e.g., mutation as 1 and wild-type as 0) and multiple logistic regression adjusting for covariates were carried out using the *glm* function in R. In pan-cancer analyses, gender, race and cancer type were variables included in the linear model. Clinical variables used in cancer-specific analyses included gender, race, pathologic stage, neoplasm histologic grade, smoking status, alcohol consumption and cancer-specific variables such as oestrogen receptor (ER) status in breast cancer (**Table 3.2**). To avoid the potential detrimental effect caused by missing data, only variables with missing data less than 10% of samples used in the somatic copy number alteration analysis were retained. To account for the difference in the proportion of cancer cells in each tumour, tumour purity (cancer cell fraction) estimated from ASCAT was included in the linear model. When necessary, to avoid the separation problem that might occur due to the sparse-data bias (Greenland *et al.*, 2016), *logistf* function from the *logistf* package (version 1.23) (Heinze & Ploner, 2018) was used to perform multivariable logistic regression with Firth's penalization (Heinze & Schemper, 2002). Effect sizes from logistic regression analyses were reported as odds ratio per year and 95% confidence intervals. *P*-values from the analyses were accounted for multiple-hypothesis testing using Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995). Statistical significance was considered if adj. p-value < 0.05, unless specifically indicated otherwise.

All statistical analyses were carried out using R (version 3.6.3) (Team, 2020). Plots were generated using *ggplot2* (version 3.3.2) (Wickham, 2016), *ggrepel* (version 0.8.2) (Slowikowski, 2020), *ggpubr* (version 0.4.0) (Kassambara, 2020a), *ComplexHeatmap* (version 2.2.0) (Gu *et al.*, 2016), and *VennDiagram* (version 1.6.20) (Chen & Boutros, 2011).

3.2.3 GI score analysis

The GI scores for TCGA cancers were obtained from our collaborators, Dr Peter Van Loo and Dr Tom Lesluyes at the Francis Crick Institute, London. GI score was calculated as a genome fraction (per cent-based) that does not fit the estimated tumour ploidy, 2 for normal diploid, and 4 for tumours that have undergone the WGD process. Simple linear regression was performed to identify the association between age and GI score. For pan-cancer analysis, multiple linear regression was used to adjust for gender, race, and cancer type. For cancer-specific analysis, multiple linear regression accounting for clinical variables was conducted on the cancer types that had a significant association between age and GI score from the simple linear regression analysis (adj. p-value < 0.05).

3.2.4 Percentage genomic LOH quantification and analysis

To quantify the per cent genomic LOH for each tumour, allele-specific copy number profiles from ASCAT were used. X and Y chromosome regions were discarded from the analysis. The LOH segments were segments that harbour only one allele. The per cent genomic LOH was defined as 100 times the total length of LOH regions / length of the genome.

Simple linear regression and multiple linear regression adjusting for gender, race, and cancer types were conducted to investigate the relationship between age and the per cent genomic LOH in the pan-cancer analysis. For cancer-specific analysis, simple linear regression was performed, followed by multiple linear regression accounting for clinical factors for cancers with a significant association in simple linear regression analysis (adj. p-value < 0.05).

3.2.5 WGD analysis

WGD status for each tumour was obtained from a fraction of the genome with LOH and tumour ploidy. To investigate the association between age and WGD across the pan-cancer dataset, simple logistic regression and multiple logistic regression correcting for gender, race, and cancer type were performed. For cancer-specific analysis, simple logistic regression was performed to access the association between age and WGD on tumours from each cancer type. Cancer types with a significant association between age and WGD (adj. p-value < 0.05) were further subjected to the multiple logistic regression accounting for the clinical variables.

3.2.6 List of known cancer driver genes

A list of known cancer driver genes was compiled from (1) the list of 243 COSMIC classic genes from COSMIC database version 91 (Tate *et al.*, 2019) (downloaded on 1st July 2020), (2) the list of 260 significantly mutated genes from Lawrence *et al.* (Lawrence *et al.*, 2014), and (3) the list of 299 cancer driver genes from the TCGA Pan-Cancer study (Bailey *et al.*, 2018). In total, the list consists of 505 cancer genes, which was used for the mutations and focal-level SCNAs analyses in this chapter.

3.2.7 Recurrent SCNA analysis

Recurrent arm-level and focal-level SCNAs of each cancer type were identified using GISTIC2.0 (Mermel *et al.*, 2011). Segmented files (nocnv_hg19.seg) from TCGA, marker file and CNV file, provided by GISTIC2.0, were used as input files. The parameters were set as follows: '-genegistic 1 -smallmem 1 -qvt 0.25 -ta 0.25 -td 0.25 -broad 1 -brlen 0.7 -conf 0.95 -armpeel 1 -savegene 1'. Based on these parameters, broad events were defined as the alterations happen in more than 70% of an arm. The log2 ratio thresholds for copy number gains and losses were 0.25 and -0.25, respectively. The confidence level was set as 0.95, and the q-value was 0.25.

To investigate the association between age and arm-level SCNAs for each cancer type, simple logistic regression was performed for each chromosomal arm that was identified as recurrent SCNA in a cancer type. Only cancer types with more than 100 samples were included in this analysis (**Table 3.1**). Arms with a significant association (adj. p-value < 0.05) were further adjusted for clinical variables using multiple logistic regression. Similarly, simple and multiple logistic regression was conducted on the focal-level SCNAs for each cancer type. Regions that are overlapped with centromeres or telomeres were removed from the analysis.

To confirm the impact of SCNAs on gene expression, the correlation between GISTIC2.0 score and RNA-seq based gene expression (log2(normalised RSEM + 1)) for tumours that have both types of data was investigated using Pearson correlation. The correlation was considered significant if the p-value corrected for multiple-hypothesis testing using the Benjamini-Hochberg procedure < 0.05.

3.2.8 SCNA score quantification and analysis

Previous studies have developed the SCNA score representing the SCNA level of a tumour (Davoli *et al.*, 2017; Yuan *et al.*, 2018). The methods described by Yuan *et al.* (Yuan *et al.*, 2018) to calculate SCNA scores were applied in the analysis. Using SCNA profiles from GISTIC2.0 analysis, SCNA scores for each tumour were derived at three different levels (chromosome-, arm-, and focal-level). For each tumour, each focal-event log2 copy number ratio from GISTIC2.0 was classified into the following score: 2 if the log2 ratio ≥ 1, 1 if the log2 ratio < 1 and ≥ 0.25, 0 if the log2 ratio < 0.25 and ≥ -0.25, -1 if the log2 ratio < -0.25 and ≥ -1, and -2 if the log2 ratio < -1. The |score| from each focal event in a tumour was then summed into a focal score of a tumour. Thereafter, the rank-based normalisation (rank/number of tumours in a cancer type) was applied to focal scores from all tumours within the same cancer type, resulting in normalised focal-level SCNA scores. Therefore, tumours with high focal-level SCNAs will have focal-level SCNA scores close to 1, while tumours with low focal-level SCNAs will have scores close to 0. For the arm- and chromosome-level SCNA scores, a similar procedure was applied to the broad event log2 copy number ratio from GISTIC2.0. An event was considered a chromosome level if both arms had the same log2 ratio; otherwise, it was considered an arm-level. Similar to the focal-level SCNA score, each arm- and chromosome-event log2 copy number ratio was classified into the 2, 1, 0, -1, -2 scores using the threshold described above. The |score| from all arm-events and chromosome-events for a tumour were then summed into an arm score and chromosome score, respectively. For each cancer type, the rank-based normalisation was applied to arm scores and chromosome scores from all tumours to derive normalised arm-level SCNA scores and normalised chromosome-level SCNA scores, respectively. An overall SCNA score for a tumour was defined as the sum of focal-level, arm-level, and chromosome-level SCNA scores. A chromosome/arm-level SCNA score for a tumour was defined as the sum of chromosome-level and arm-level SCNA scores.

The association between age and overall, chromosome/arm-level, and focal-level SCNA scores for each cancer type was investigated using simple linear regression. Cancer types with a significant association (adj. p-value < 0.05) were then subjected to multiple linear regression analysis adjusting for the clinical variables.

3.2.9 Analysis of age-associated somatic mutation in cancer genes

Somatic mutation data was obtained from the MAF file from the recent TCGA MC3 project (Ellrott *et al.*, 2018). In the MC3 effort, variants were called using seven variant callers to enable robust variant calling. All varients were filtered to keep only non-silent SNVs and indels located in gene bodies, retaining only "Frame_Shift_Del", "Frame_Shift_Ins", "In_Frame_Del", "In_Frame_Ins", "Missense_Mutation", "Nonsense_Mutation", "Nonstop_Mutation", "Splice_Site" and "Translation_Start_Site" in the "Variant_Classification" column. I focused only on mutations in the cancer genes from the compiled list of cancer driver genes. To prevent the bias that might cause by hypermutated tumours, the analysis was restricted to tumours with < 1,000 mutations per exome. For pan-cancer analysis, multiple logistic regression accounting for gender, race and cancer type was performed to investigate the association between age and mutations in 20 cancer genes that are mutated in > 5% of samples. For cancer-specific analysis, simple logistic regression was used to identify cancer genes that the mutations in these genes are associated with the patient's age. Only genes that are mutated in > 5% of samples from each cancer type were included in the analysis. The significant associations (adj. *p*-value < 0.05) were further investigated using multiple logistic regression accounting for clinical variables.

3.2.10 Analysis of mutation burden, substitution classes, MSI-H status, and *POLE*/*POLD1* mutations

A mutation burden was defined as the total non-silent mutations in an exome. A package *maftools* (version 3.3.2) was used to import and extract information from MAF files (Mayakonda *et al.*, 2018). The mutation burden for each tumour was log-transformed before using it in the subsequent analysis. To investigate the relationship between age and mutation burden in pan-cancer, multiple linear regression adjusting for gender, race and cancer type was conducted. For cancer-specific analysis, simple linear regression was performed. Cancer types with a significant association between age and mutation burden in simple linear regression analysis (adj. *p*-value < 0.05) were further examined using multiple linear regression accounting for clinical factors. Similarly, multiple linear regression adjusting for gender, race and cancer type was used to examine the relationship between age and each of the six substitution classes in pan-cancer. For cancer-specific analysis, simple linear regression was performed. Cancer types with a significant association between age and fraction contribution

of a substitution class in simple linear regression analysis (adj. *p*-value < 0.05) were further examined using multiple linear regression accounting for clinical factors.

Microsatellite instability status for COAD, READ, STAD, and UCEC were downloaded from TCGA using *TCGAbiolinks*. To study the association between the presence of high microsatellite instability (MSI-H) and age, tumours were divided into binary groups: MSI-H = TRUE and MSI-H = FALSE. Multiple logistic regression adjusting for clinical variables was then performed. Similarly, *POLE* and *POLD1* mutation status were in a binary outcome (mutated and not mutated). Multiple logistic regression was used to investigate the association between age and *POLE*/*POLD1* mutations in cancer types that contained POLE/POLD1 mutations in > 5% of samples.

3.2.11 Oncogenic signalling pathway analysis

The list of pathway-level alterations in ten oncogenic pathways (cell cycle, Hippo, Myc, Notch, Nrf2, PI-3-Kinase/Akt, RTK-RAS, TGFβ signalling, p53 and β-catenin/Wnt) for TCGA tumours comprehensively compiled by Sanchez-Vega *et al.* (Sanchez-Vega *et al.*, 2018) was used in this analysis. Member genes in the pathways were accessed for SCNAs, mutations, epigenetic silencing through promoter DNA hypermethylation and gene fusions. Only the pathway alteration data of samples that were presented in the SCNA analysis was retained. For the pan-cancer analysis, multiple logistic regression adjusting for the patient's gender, race and cancer type was used to demonstrate the relationship between pathway-level alteration and age. To investigate the association between age and cancer-specific pathway alterations, simple logistic regression was performed. Cancer types with a significant association (adj. *p*-value < 0.05) were further examined by multiple logistic regression accounting for clinical variables.

3.2.12 Gene expression and DNA methylation analysis

To render the results from gene expression and DNA methylation comparable, the analysis was limited to genes that are presented in both types of data. The lowly expressed genes were filtered out from the analysis by keeping only genes with RSEM > 0 in more than 50 per cent of samples. Only protein-coding genes identified using biomaRt (Durinck *et al.*, 2009) (version 2.46.0, data based on Ensembl version 100, April 2020) were included in the analyses. Normalised mRNA expression in RSEM for each TCGA cancer type was log2-transformed before being subjected to the multiple linear regression analysis adjusting for clinical factors. RNA-seq data for colon cancer and endometrial cancer consisted of two platforms, Illumina HiSeq and Illumina GA. Thus, a platform was included as another covariate in the linear regression model for these two cancer types. Genes with adj. *p*-value < 0.05 were considered significantly differentially expressed genes with age (age-DEGs). DNA methylation data was presented as β-values, which are the ratio of the intensities of methylated and unmethylated alleles. Because multiple methylation probes can be mapped to the same gene, the one-to-one mapping genes and probes was done by selecting the probes that are most negatively correlated with the corresponding gene expression from the meth.by_min_expr_corr.data.txt files downloaded from Broad GDAC Firehose. Similar multiple linear regression to the gene expression analysis was performed on the methylation data. Genes with adj. *p*-value < 0.05 were considered significant differentially methylated genes with age (age-DMGs). To exclude the possibility that germline predisposition mutations in some patients may cause this high number of age-DEGs and age-DMGs in female reproductive cancers, samples harbouring germline mutations in *BRCA1*, *BRCA2* and *TP53* as previously identified (Huang *et al.*, 2018) were excluded from the analysis of breast, ovarian and endometrial cancer cohorts. In total, ~13% of patients from breast and endometrial cancer and ~20% of patients from ovarian cancer were removed. After excluding these samples, a similar multiple linear regression analysis was performed.

The correlation between gene expression and DNA methylation was calculated using Pearson correlation. The Kruskal-Wallis test was used to investigate the differences between correlation coefficients among groups (age-DMGs-DEGs, age-DMGs, age-DEGs, other genes). The pairwise comparisons were carried out by a two-sided Dunn's test.

Gene Set Enrichment Analysis (GSEA) was performed to investigate the Gene Ontology (GO) terms that are enriched in tumours from younger or older patients. The analysis was done using the package *ClusterProfiler* (version 3.14.3) (Yu *et al.*, 2012). Briefly, genes or methylation probes were ranked based on their regression coefficient with age from the most positive regression coefficient with age (most up-regulated in older patients) to the gene with the most negative regression coefficient with age (most up-regulated in younger patients). A ranked gene list consisting of all genes was used in the GSEA to determine whether genes in a set of interest from GO are randomly distributed throughout the ranked gene list or are found more often in the top or the bottom of the list (Subramanian *et al.*, 2005).

3.3 Results

3.3.1 Association between age and genomic instability, loss of heterozygosity, and whole-genome duplication

To gain insight into the role of patient age into the somatic genetic profile of tumours, I evaluated associations between patient age and genomic features of tumours in TCGA data (**Table 3.1**). I first obtained genomic instability (GI) scores, which were calculated as the genome fraction (percent-based) that does not fit the ground state, defined as 2 for tumours that have not undergone whole-genome duplication (WGD), and 4 for tumours that have undergone WGD (Methods). Using multiple linear regression adjusting for gender, race, and cancer type, I found that GI scores increase with age in pan-cancer data (adj. R-squared = 0.35, *p*-value = 5.98×10^{-7}) (**Figure. 3.1A**). I next applied simple linear regression to investigate the relationship between GI scores and age for each cancer type. Cancer types with a significant association (adj. *p*-value < 0.05) were further adjusted for clinical variables. For this cancer-specific analysis, I found a significant positive association between age and GI score in seven cancer types (adj. *p*-value < 0.05) (**Figure. 3.1B**). Cancer types with the strongest significant positive association were low-grade glioma, ovarian cancer, endometrial cancer and sarcoma (**Figure. 3.1C**). This result indicates that the level of genomic instability increases with the age of cancer patients in several cancer types, consistent with the finding from a recent study (Li *et al.*, 2022).

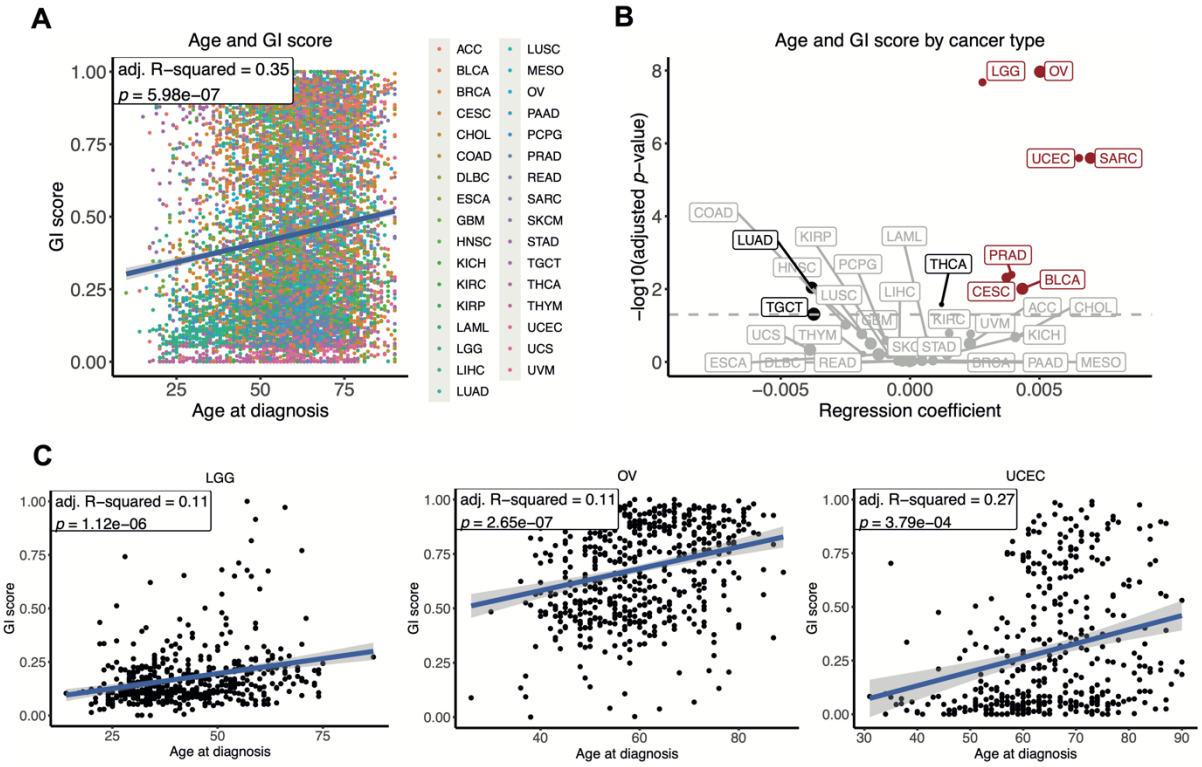


Figure 3.1. Association between cancer patient's age and genomic instability (GI) score. **(A)** Association between age and pan-cancer GI score. Dots are coloured by cancer type. Multiple linear regression R-squared and p-value are shown in the figure. **(B)** Association between age and cancer type-specific GI score. Linear regression coefficients and significant values are shown in the figure. Cancers with a significant positive association between age and GI score after using multiple linear regression (adj. *p*-value < 0.05) are highlighted in red. Cancers with a significant association in simple linear regression but not significant after using multiple linear regression are shown in black. The grey line indicates adj. *p*-value = 0.05. Dot size is proportional to the median GI score. **(C)** Example of the association between age and GI score in low-grade glioma, ovarian cancer, and endometrial cancer.

Genomic loss of heterozygosity (LOH) refers to the irreversible loss of one parental allele, causing an allelic imbalance, and priming the cell for another defect at the other remaining allele of the respective genes (Ryland *et al.*, 2015). To investigate whether there is an association between patients' age and LOH, I quantified per cent genomic LOH from the allele-specific copy number data. The LOH segments were segments that harbour only one allele. By using simple linear regression, I found a significant positive association between age and pan-cancer per cent genomic LOH (p-value = 1.20×10^{-21}). However, this association was no longer significant in a multiple linear regression analysis (adj. R-squared = 0.32, p-value = 0.289) (**Figure 3.2A**). Thus, it is likely that this association might be cancer type-specific. I then performed a linear regression between age and per cent genomic LOH for each cancer type. Six cancer types showed a positive association between age and per cent genomic LOH (adj. p-value < 0.05) (**Figure 3.2B**). The strongest positive associations were found in low-grade glioma and endometrial cancer (adj. p-value < 0.05) (**Figure 3.2C**), corroborating with the increase in GI score with age. On the other hand, lung adenocarcinoma (**Figure 3.2C**), oesophageal and liver cancer demonstrated a negative correlation between per cent genomic LOH and age (adj. p-value < 0.05). This negative correlation might be due to the difference in the distribution of age of samples with smoking status (lung adenocarcinoma and oesophageal cancer), race (oesophageal cancer) and tumour grade (liver cancer) (**Figur 3.3**), yet other unexplained factors might also contribute to the higher LOH level in younger patients in these three cancer types.

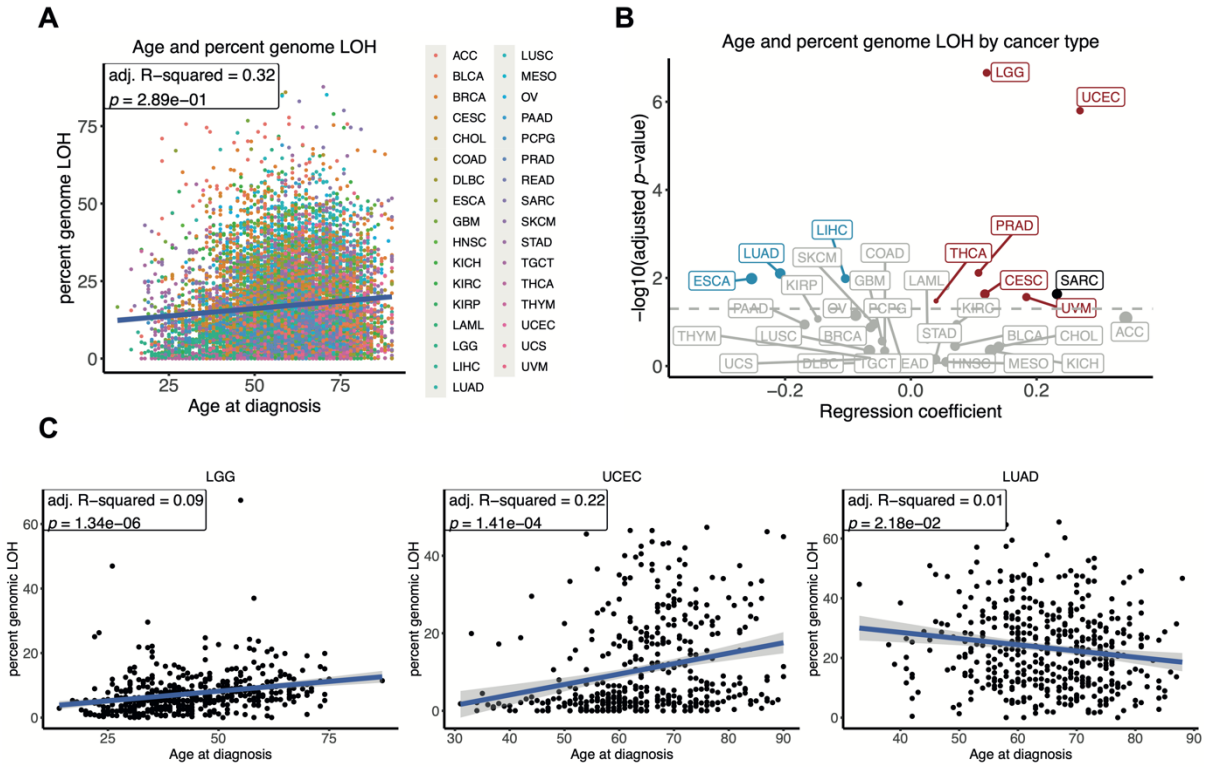


Figure 3.2. Association between cancer patients' age and percent genomic loss-of-heterozygosity (LOH).

(A) Association between age and pan-cancer percent genomic LOH. Dots are coloured by cancer type. Multiple linear regression R-squared and *p*-value are shown in the figure. **(B)** Association between age and cancer type-specific percent genomic LOH. Simple linear regression results are shown. Cancers with a significant positive and negative association between age and percent genomic LOH after using multiple linear regression are highlighted in red and blue, respectively. Cancer with a significant association in simple linear regression but not significant after using multiple linear regression is showed in black. The grey line indicates adj. *p*-value = 0.05. Dot size is proportional to median percent genomic LOH. **(C)** Example of the association between age and LOH in low-grade glioma, endometrial cancer, and lung adenocarcinoma.

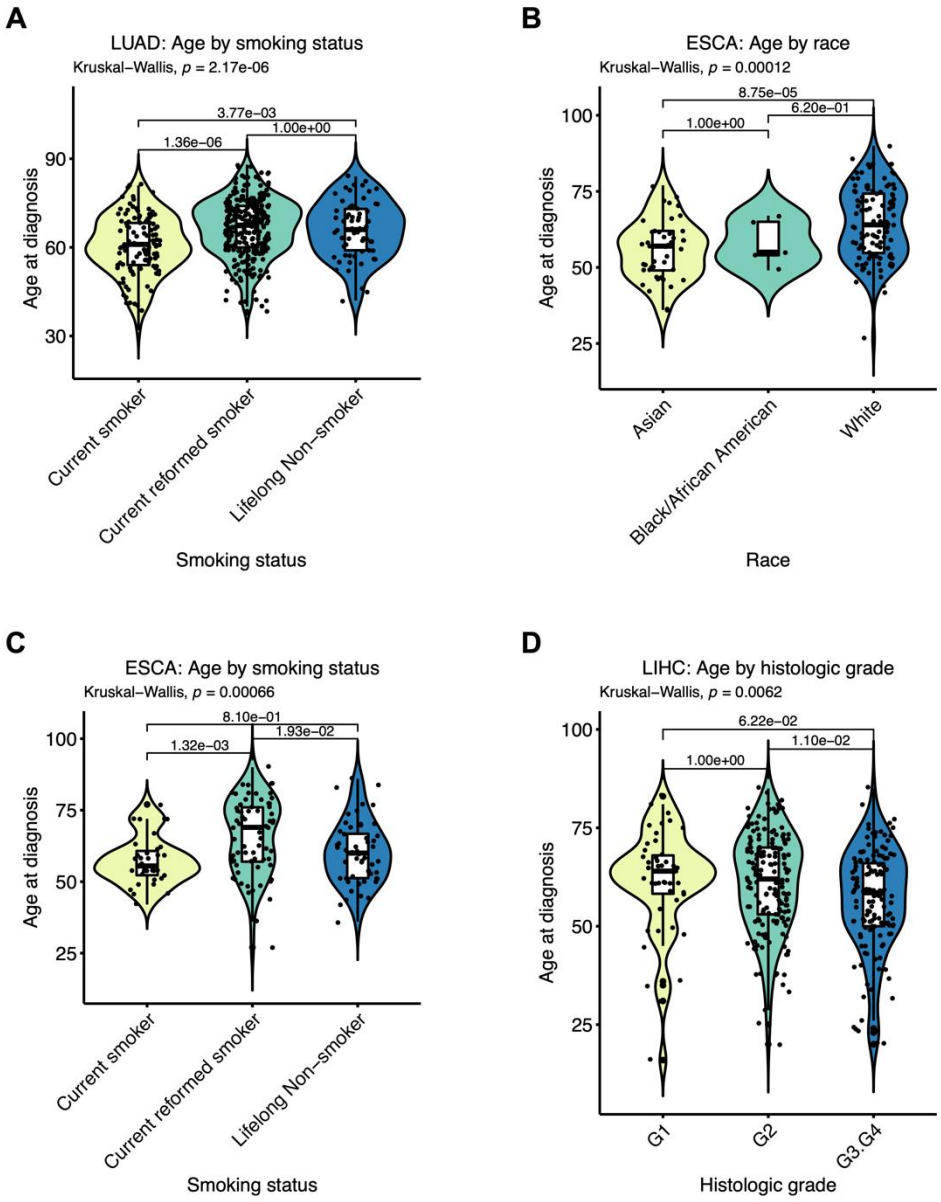


Figure 3.3. Age distribution of patients presented with different clinical factors.

(A) Age distribution of lung adenocarcinoma patients by smoking status. Current smoker *n* = 112, Current reformed smoker *n* = 276, Lifelong Non-smoker *n* = 62 samples. **(B)** Age distribution of oesophageal cancer patients by race. Asian *n* = 44, Black/African American *n* = 5, White *n* = 10 samples. **(C)** Age distribution of oesophageal cancer patients by smoking status. Current smoker *n* = 34, Current reformed smoker *n* = 70, Lifelong Non-smoker *n* = 54 samples. **(D)** Age distribution of liver cancer patients by histologic grade. G1 *n* = 46, G2 *n* = 170, G3 and G4 *n* = 134 samples. The group comparison was performed by the Kruskal-Wallis test. The pairwise comparisons were done using the two-sided Dunn's test. The middle bar of the boxplot is the median. The box represented interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of 1.5 x IQR.

WGD is important in increasing the adaptive potential of the tumour by being a buffer against the adverse effects of deleterious mutations or copy-number alterations (Dewhurst *et al.*, 2014). WGD has also been linked with a poor prognosis (Van de Peer *et al.*, 2017; Bielski *et al.*, 2018; Lopez *et al.*, 2020). I, thus, investigated the relationship between age and WGD events using logistic regression. In a pan-cancer analysis, I found a slight increase in the probability that WGD occurs with age, using multiple logistic regression accounting for gender, race, and cancer type (odds ratio per year (OR) = 1.0066, 95% confidence interval (CI) = 1.0030-1.0103, *p*-value = 3.84 x 10^{-4}) (**Figure 3.4A**). For the cancer-specific analysis, a significant positive association was found in ovarian and endometrial cancer (OR = 1.0320 and 1.0248, 95%CI = 1.0151-1.0496 and 1.0024-1.0483, adj. *p*-value = 4.68x10^{-4} and 0.049, respectively) (**Figure 3.4B-C**), indicating that tumours from older patients are more likely to have doubled their genome.

Taken together, the findings indicate that tumours from patients with an increased age tend to harbour a more unstable genome and a higher level of LOH in several cancer types. Notably, the strongest association between age and an increase in genome instability, LOH, and WGD was evident in endometrial cancer, suggesting potential disparities in the cancer genome landscape with age in this cancer type.

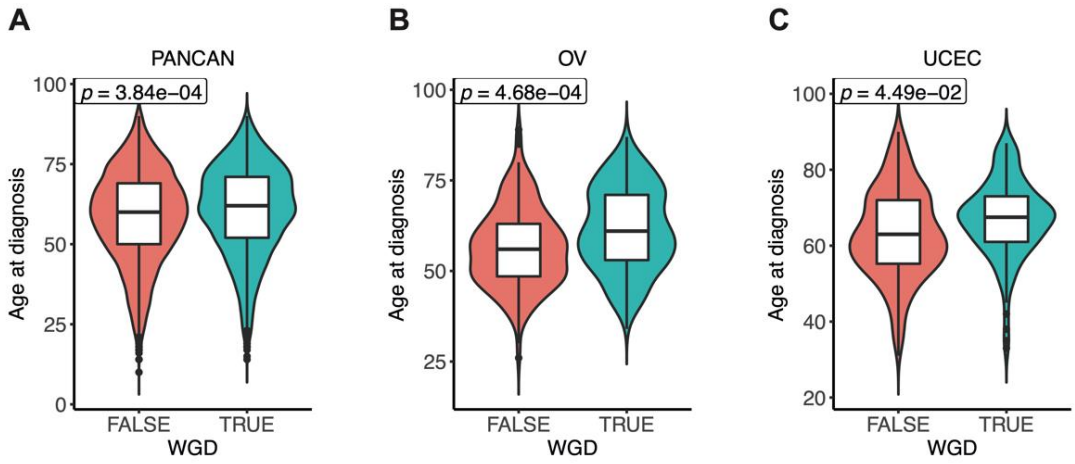


Figure 3.4. Association between cancer patient's age and whole-genome duplication events (WGD).

Association between age and WGD events in **(A)** pan-cancer, **(B)** ovarian cancer, and **(C)** endometrial cancer. Multiple logistic regression *p*-values were indicated in the figure. The middle bar of the boxplot is the median. The box represents the interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of 1.5 x IQR.

3.3.2 Age-associated somatic copy-number alterations

I used GISTIC2.0 to identify recurrently altered focal- and arm-level SCNAs (Mermel *et al.*, 2011). I calculated the SCNA score as a representation of the level of SCNA occurring in a tumour (Davoli *et al.*, 2017; Yuan *et al.*, 2018). For each tumour, the SCNA score was calculated at three different levels: focal-, arm- and chromosome-level, and the overall score calculated from the sum of all three levels. I used simple linear regression to identify the association between age and overall SCNA scores. Cancer types that displayed a significant association were further adjusted for clinical variables. Consistent with the GI score results described above, the strongest positive association between age and overall SCNA scores was found in low-grade glioma, ovarian and endometrial cancers. Other cancer types for which a positive association between age and overall SCNA score was observed were thyroid cancer and clear cell renal cell carcinoma (adj. *p*-value < 0.05). On the other hand, lung adenocarcinoma is the only cancer type exhibiting a negative association between overall SCNA score and age (**Figure 3.5A**), possibly due to the presence of current smokers in younger

lung adenocarcinoma patients (**Figure 3.3A**). When I analysed only non-smokers, there was no significant association between age and overall SCNA score. However, a significant negative association was found when I analysed only current reformed smokers and only current smokers, thus, other unexplained factors apart from smoking status might also contribute to this higher SCNA score in younger smokers.

The different SCNA classes (focal- and chromosome/arm-level) may arise through different biological mechanisms (Van de Peer *et al.*, 2017; Yuan *et al.*, 2018). Therefore, I separately analysed the association between age and focal- and chromosome/arm-level SCNA scores. Most cancer types that showed a significant relationship between age and overall SCNA score also had an association between age and both chromosome/arm-level and focal-level SCNA scores (**Figure 3.5B-C**). The only exception was sarcoma, with a significant association between age and chromosome/arm-level but not with focal-level and overall SCNA scores.

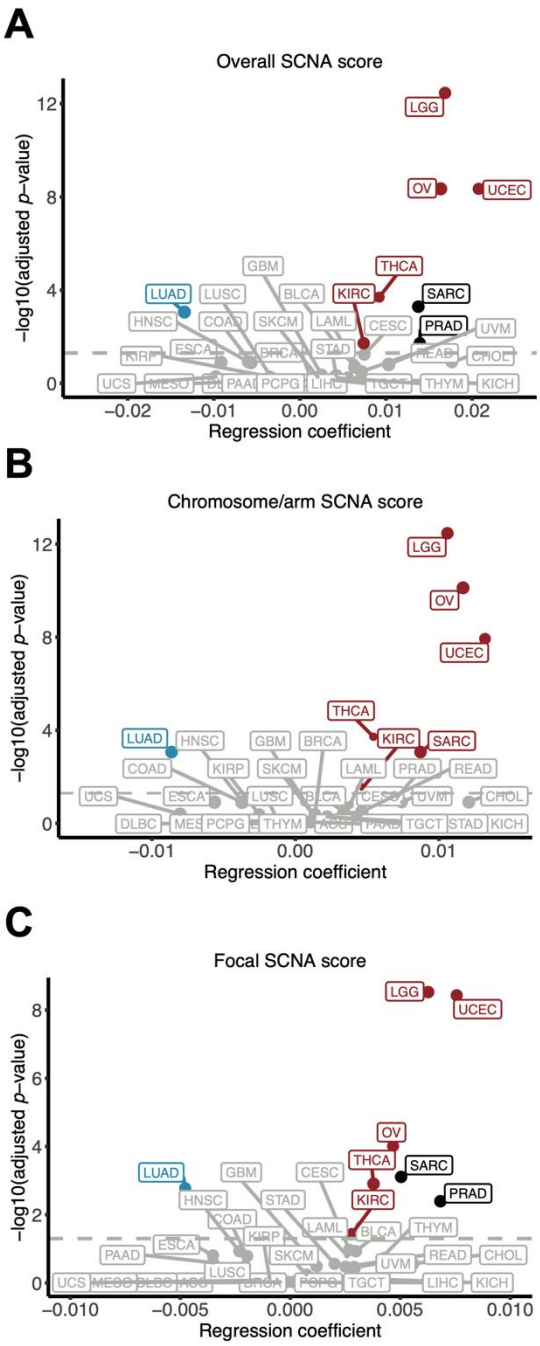


Figure 3.5. Association between cancer patients' age and somatic copy-number alterations (SCNA) scores.

Volcano plot representing the association between age and **(A)** overall, **(B)** focal-level and **(C)** chromosome/arm-level SCNA scores. Linear regression coefficients and significant values are shown. Cancers with a significant positive and negative association between age and SCNA score after using multiple linear regression (adj. *p*-value < 0.05) are highlighted in red and blue, respectively. Cancers with a significant association in simple linear regression but not significant after using multiple linear regression are shown in black. The grey line indicates adj. *p*-value = 0.05. Dot size is proportional to the median SCNA score.

I next identified the chromosomal arms that tend to be gained and lost more often with age for 25 cancer types with sufficient samples (at least 100 tumours, **Table 3.1**). I conducted logistic regression on the significant recurrently gained and lost arms that were identified by GISTIC2.0 for each cancer type. The significant associations between age and chromosomal arm gains and losses are shown in **Figure 3.6A-B**, respectively (adj. *p*-value < 0.05). Gains of chromosome 7p, 7q, 20p, and 20q significantly increased with age in several cancer types, including two types of gliomas, low-grade glioma and glioblastoma. On the other hand, the gain of chromosome 10p decreased with increased age in gliomas (**Figure 3.6A and Figure 3.7A**). For the arm losses, there was an increased occurrence of loss in 11 arms with advanced age in endometrial cancer (**Figure 3.6B and Figure 3.7B**), consistent with a higher genomic instability and LOH with age in this cancer type. These arms included 9p and 17p, containing tumour suppressor genes *CDKN2A* and *TP53*, respectively. Low-grade glioma and ovarian cancer, two other cancer types for which I found the highest significant association between age and SCNA scores, also exhibited a significant increase or decrease in losses with age in multiple arms (**Figure 3.6B and Figure 3.7A**). I also observed that losses of chromosomes 10p and 10q increased with age in gliomas. Recurrent losses of chromosome 10 (containing *PTEN*), together with gains of chromosome 7 (containing *EGFR*) are important features in IDH-wild-type (IDH-WT) gliomas (Korber *et al.*, 2019). This type of glioma was more common in older patients, whereas IDH-mutant gliomas were predominantly found in younger patients. Apart from gliomas and endometrial cancer, arm-level gains and losses in other cancer types are also related to known cancer-driver genes. For instance, I found an increased incidence in the loss of chromosome 13q (harbouring *RB1*) with age in thyroid cancer. Gains of chromosome 12 (containing the *KRAS* oncogene) increased with age in ovarian cancer. Indeed, while I can explain some of the age-associated chromosomal arm alterations, further closer inspection of arm-level alterations is required to fully explain why some specific arms are more or less frequently gained or lost as a function of age in particular cancer types.

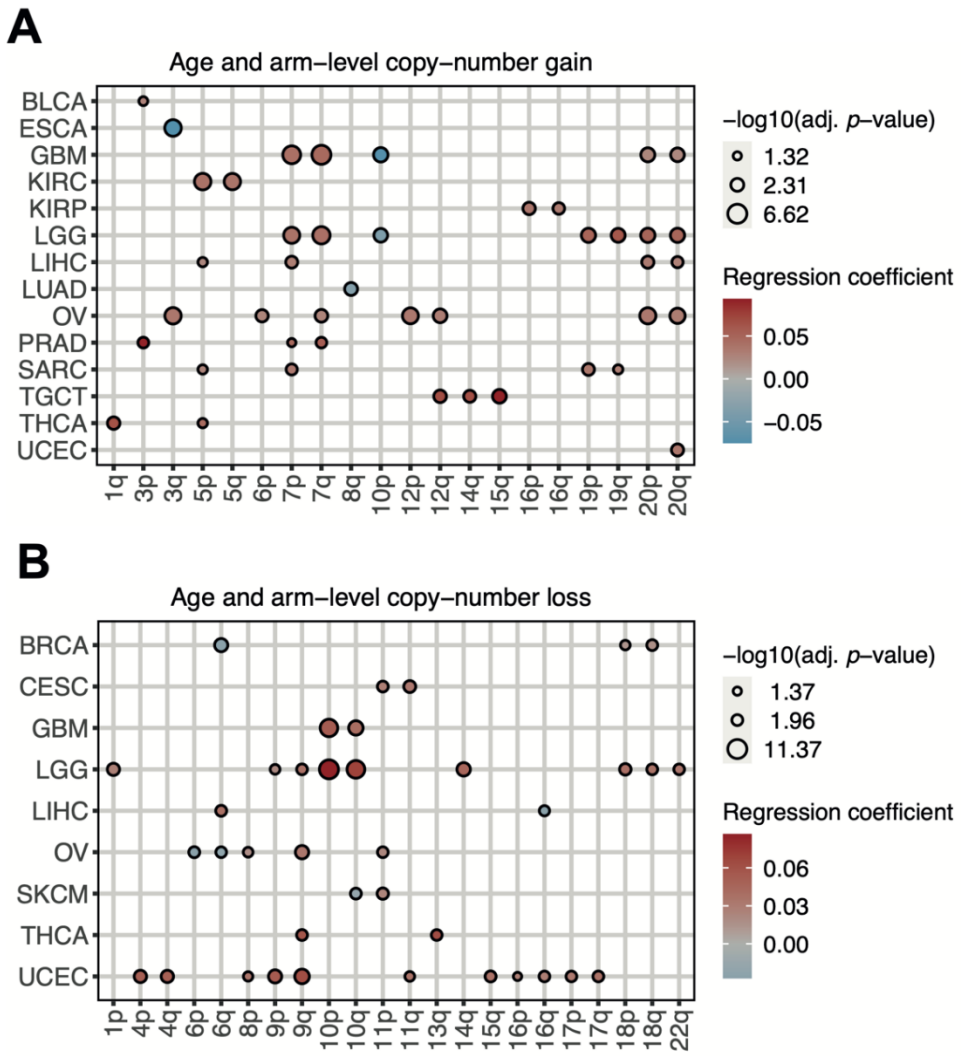


Figure 3.6. Association between cancer patients' age and somatic copy-number alterations (SCNAs).

The dot plots show the association between age and **(A)** arm-level copy-number gains and **(B)** arm-level copy-number losses. Circle size corresponds to the significant level; red and blue represent positive and negative associations, respectively.

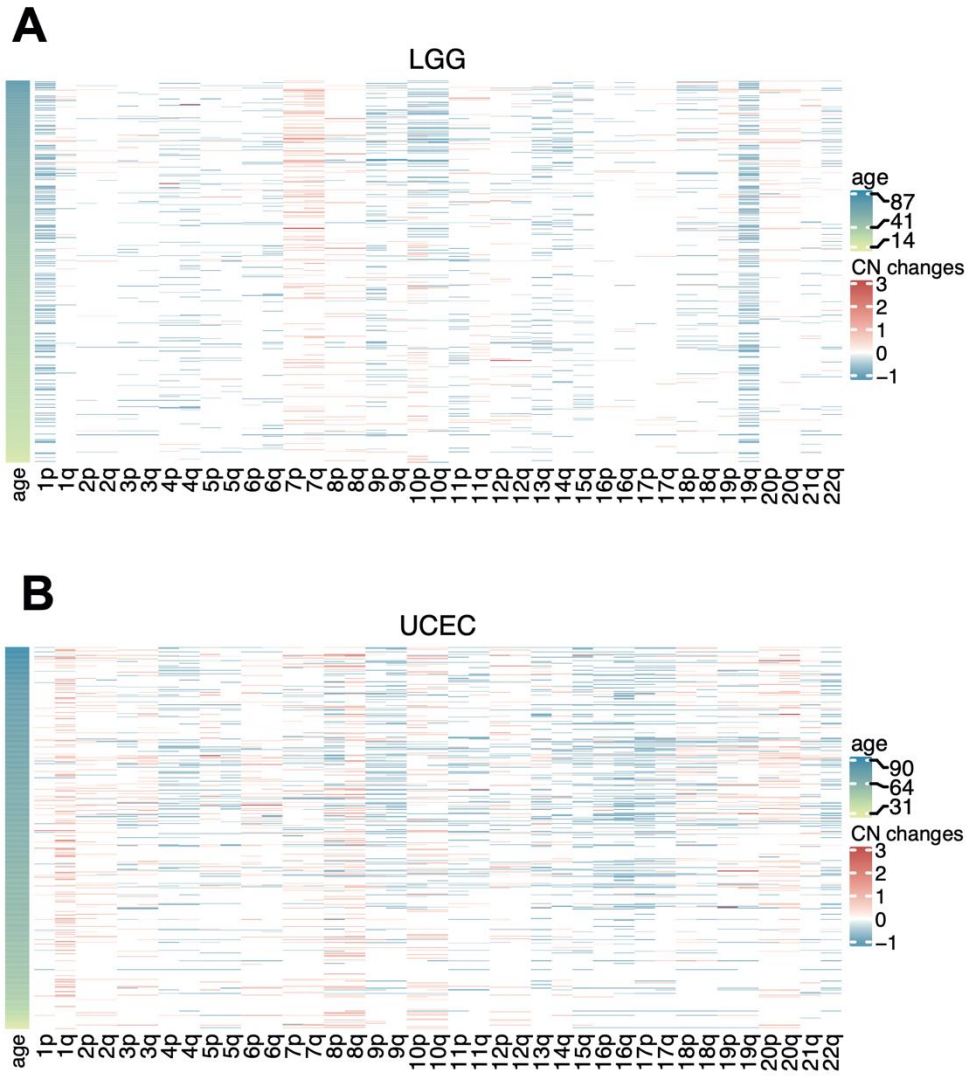


Figure 3.7. Examples of arm-level copy-number alterations.

Heatmaps represent arm-level copy-number alterations in **(A)** low-grade glioma and **(B)** endometrial cancer, respectively. Samples are sorted by age. Colours represent copy-number changes from GISTIC2.0; blue denotes loss, and red corresponds to gain.

I further examined age-associated recurrent focal SCNAs. Applying a similar logistic regression, I identified recurrent focal SCNAs associated with the age of the patients for each cancer type. In total, I found 113 significant age-associated regions, including 67 gains across 10 cancer types and 46 losses across 9 cancer types (adj. *p*-value < 0.05) (**Figure 3.8A-B**). In accordance with the arm-level result, the highest number of significant regions were found in endometrial cancer (23 gains and 25 losses), followed by ovarian cancer (13 gains and 2 losses) and low-grade glioma (9 gains and 5 losses) (**Figure 3.9**).

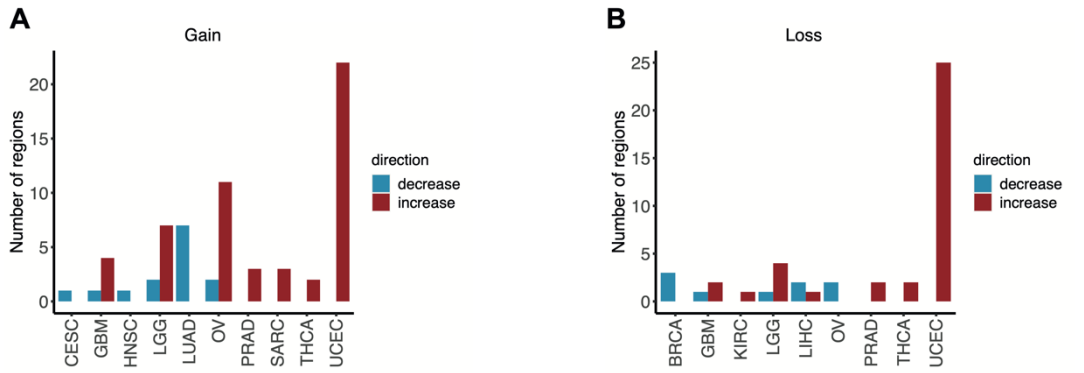


Figure 3.8. Association between cancer patient's age and focal-level SCNAs.

The number of focal regions with **(A)** copy-number gains and **(B)** copy-number losses that showed a significant association with age per cancer type (multiple logistic regression, adj. *p*-value < 0.05).

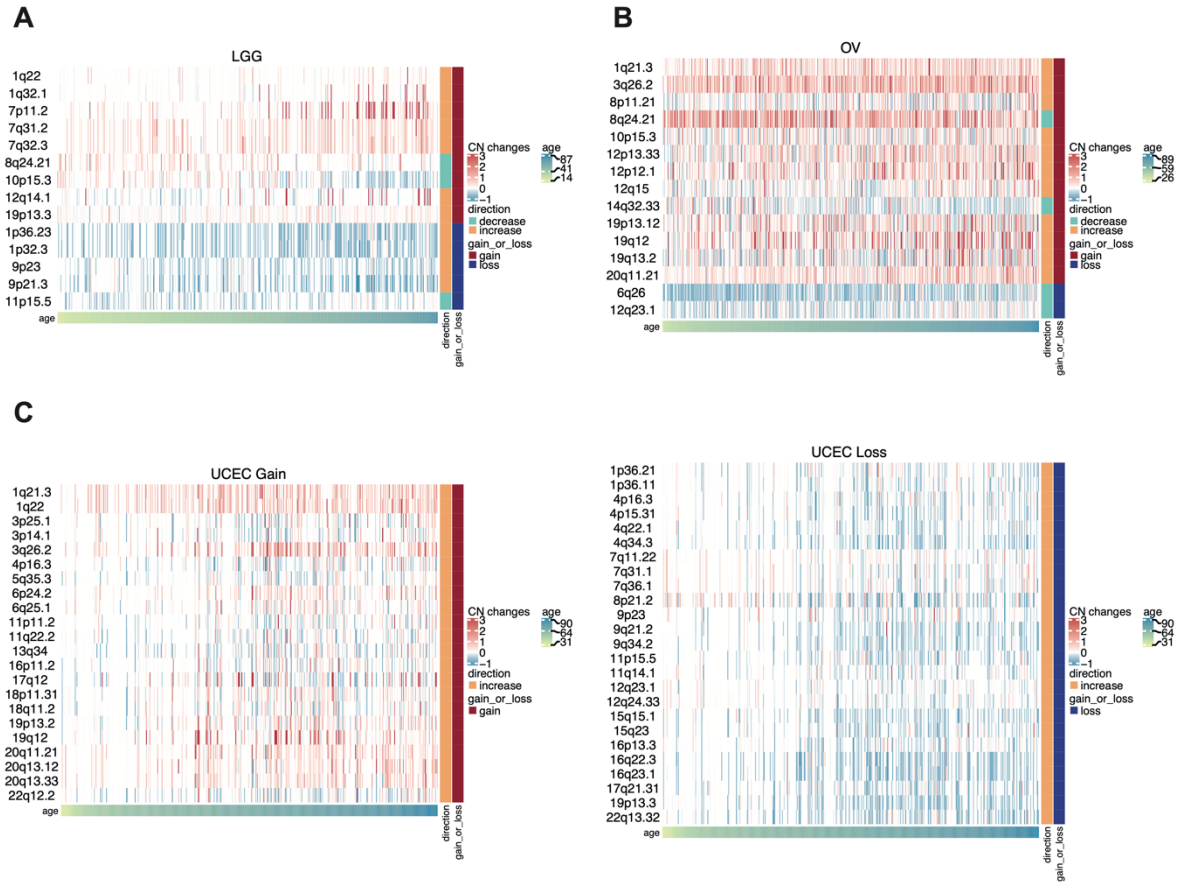


Figure 3.9. Examples of age-associated focal-level copy-number alterations.

Heatmap showing age-associated focal-level SCNAs in **(A)** low-grade glioma and **(B)** ovarian cancer, and **(C)** endometrial cancer. Samples are sorted by age. Colours represent copy-number changes from GISTIC2.0; blue denotes loss, and red corresponds to gain. The gain_or_loss legend demonstrates that the region is recurrently gained or deleted. The direction legend shows whether the gain/loss of the region increases or decreases with age.

To further investigate the impact of these SCNAs, I studied the correlation between the SCNAs and gene expression for tumours that have both types of data using Pearson correlation. In total, 81 genes in the list of previously identified cancer driver genes (Lawrence *et al.*, 2014; Bailey *et al.*, 2018; Tate *et al.*, 2019) were presented in at least one significant age-associated focal region in at least one cancer type and showed a significant correlation between SCNA and gene expression (adj. *p*-value < 0.05) (**Figure 3.10A**). For example, regions showing an increased gain with age in endometrial cancer included 1q22, where the gene *RIT1* is located in (OR = 1.0355, 95%CI = 1.0151-1.0571, adj. *p*-value = 0.0018) (**Figure 3.9C and Figure 3.10**). The Ras-related GTPases *RIT1* has been reported to be highly amplified and correlated with poor survival in endometrial cancer (Xu *et al.*, 2015). Therefore, an increased incidence of *RIT1* gains with age might relate to a poor prognosis in older patients. The 16p13.3 loss increased in frequency in older endometrial cancer patients (OR = 1.0335, 95%CI = 1.0048-1.0640, adj. *p*-value = 0.0328). This region contains the p53 coactivator gene *CREBBP*. The gain of 8q24.21 (harbouring the oncogene *MYC*) decreased with patient age in low-grade glioma (OR = 0.9737, 95%CI = 0.9541-0.9927, adj. *p*-value = 0.0128) and ovarian cancer (OR = 0.9729, 95%CI = 0.9553-0.9904, adj. *p*-value = 0.0063) (**Figure 3.10A-B**). In addition, in low-grade glioma, I found an increased incidence of 9p21.3 loss with age (OR = 1.0332, 95%CI = 1.0174-1.0496, adj. *p*-value = 0.00017). This region contains the cell cycle-regulator genes *CNKN2A* and *CDKN2B* (**Figure 3.9A and Figure 3.10**). Taken together, this analysis demonstrates the association between age and SCNAs across cancer types. I also identified age-associated arm-level and focal regions, and these regions harboured several known cancer-driver genes. These results suggest a possible contribution of different SCNA events in cancer initiation and progression of patients of different ages.

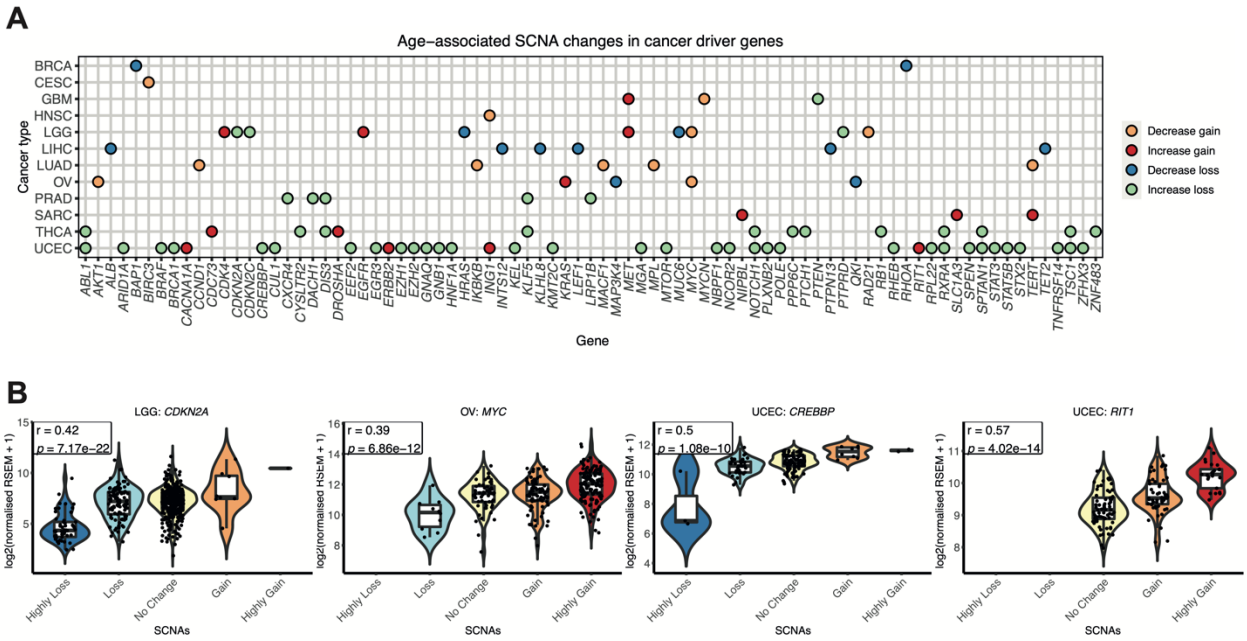


Figure 3.10. Cancer driver genes in the age-associated focal-level SCNAs.

(A) Age-associated SCNA changes in cancer driver genes. Cancer driver genes located in the age-associated focal regions are plotted by cancer type. Colours of the dot represent the condition of the focal region where the gene is located as follows: blue - decrease loss; green - increase loss; yellow - decrease gain; and red - increase gain with age. **(B)** The effect of copy-number changes on gene expression of *CDKN2A* in LGG, *MYC* in OV, *CREBBP* and *RIT1* in UCEC. These are examples of genes with age-associated changes in SCNAs. Violin plots show the log2(normalized expression + 1) of samples grouped by their SCNA status. Pearson correlation coefficient r and p-value are shown in the figures. The middle bar of the boxplot is the median. The box represents the interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of 1.5 x IQR.

3.3.3 Age-associated somatic mutations in cancer

The increase in mutation burden with age is well-established (Tomasetti *et al.*, 2013; Alexandrov *et al.*, 2015; Milholland *et al.*, 2015). This age-related mutation accumulation is in part explained by a clock-like mutational process, spontaneous deamination of 5-methylcytosine to thymine (Alexandrov *et al.*, 2015). As expected, I confirmed the positive association between age and mutation load (somatic non-silent SNVs and indels) in the pan-cancer cohort using multiple linear regression adjusting for gender, race, and cancer type (adj. R-squared = 0.53, *p*-value = 1.41 x 10^{-37}) (**Figure 3.11A**). In cancer-specific analyses, 18 cancer types exhibited a significant relationship between age and mutation load using linear regression (adj. *p*-value < 0.05) (Supplementary Figure 7a, Supplementary Data 9). This increase in mutation load was mainly contributed by C>T mutations, as I found a positive association between the fraction of C>T mutations and age (regression coefficient = 0.058, *p*-value = 8.57x10^{-7}) in a pan-cancer analysis. Conversely, the fraction of C>A mutations was negatively associated with age (regression coefficient = -0.065, *p*-value = 8.84x10^{-10}), concordant with a previous report (Milholland *et al.*, 2015). I also examined, for each cancer type, the association between age and fraction contribution of each substitution class. Consistent with the pan-cancer analysis, C>T mutations showed a significant positive association with age in six cancer types, whereas C>A mutations had a significant negative association with age in three cancer types. (**Figure 3.11B**).

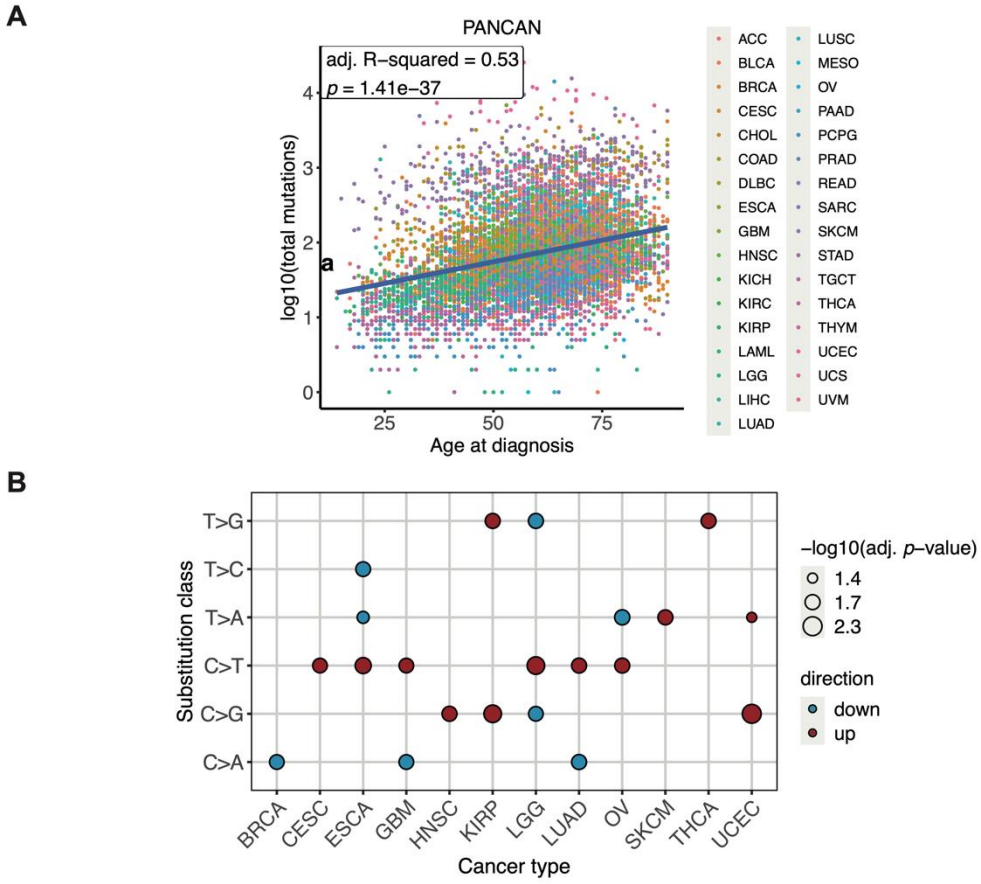


Figure 3.11. Association between cancer patients' age and somatic mutation burden.

(A) Association between patient's age and mutation burden in the pan-cancer analysis. Adjusted R-squared and p-value from multiple linear regression analysis are presented. **(B)** Association between age and fraction contribution of each substitution class. For each substitution class in each cancer, simple linear regression was performed to investigate the association between age and fraction contribution. Cancer types with a significant association (adj. p-value < 0.05) were further investigated using multiple linear regression. Multiple hypothesis testing correction was done using Benjamini–Hochberg procedure. Only significant associations from multiple linear regression analysis were plotted. Circle size corresponds to the significant level; red and blue represent positive and negative associations, respectively.

Only endometrial cancer showed a negative correlation between mutation burden and age (**Figure 3.12A**). I observed a high proportion of hypermutated tumours (> 1,000 non-silent mutations per exome) from younger endometrial cancer patients. Thirteen out of 38 tumours (34%) from the younger patients (age ≤ 50) were hypermutated tumours, while there were only 42 hypermutated tumours among the 383 tumours from older patients (11%) (two-sided Fisher's exact, *p*-value = 0.0003) (**Figure 3.13A**). Microsatellite instability (MSI) is a unique molecular alteration caused by defects in DNA mismatch repair (Kim *et al.*, 2013a; Bonneville

et al., 2017). The MSI-high (MSI-H) tumours occur as a subset of high mutation burden tumours (Chalmers *et al.*, 2017). I investigated whether high mutation loads in endometrial cancer from young patients were due to the presence of MSI-H tumours. Using multiple logistic regression, I found that MSI-H tumours were associated with younger endometrial cancer (OR = 0.9751, 95%CI = 0.9531-0.9971, *p*-value = 0.0264) (**Figure 3.13B**). Another source of hypermutation in cancer is defective DNA polymerase proofreading due to mutations in polymerase ε (*POLE*) or polymerase δ (*POLD1*) genes (Shlien *et al.*, 2015; Campbell *et al.*, 2017). I showed that mutations in *POLE* (OR = 0.9690, 95%CI = 0.9422-0.9959, *p*-value = 0.0243) and *POLD1* (OR = 0.9573, 95%CI = 0.9223-0.9925, *p*-value = 0.0177) were both more prevalent in younger endometrial cancer patients (**Figure 3.13C**). Indeed, when I excluded tumours with MSI-H and tumours containing *POLE/POLD1* mutations from the analysis, I found a significant positive association between mutation burden and age in endometrial cancer (adj. R-squared = 0.12, *p*-value = 0.00138) (**Figure 3.12B**). Therefore, the negative correlation between age and mutation loads in endometrial cancer could be explained by the presence of hypermutated tumours in younger patients, which are associated with MSI-H and *POLE/POLD1* mutations. Previous studies on *POLE* and MSI-H subtypes in hypermutated endometrial tumours revealed that these subtypes associated with a better prognosis when compared with the copy-number high subtype (Cancer Genome Atlas Research *et al.*, 2013a; Berger *et al.*, 2018; Ashley *et al.*, 2019). Together with the SCNA results, younger endometrial cancer patients are likely to associate with a *POLE* and MSI-H subtypes, high mutation rate and better survival, whilst tumours from older patients are characterized by many SCNAs and are generally associated with a worse prognosis, indicating differences between age-related subtypes in endometrial cancer. I extended the age and MSI-H analysis to other cancer types known to have a high prevalence of MSI-H tumours, including colon, rectal, and stomach cancers (Bonneville *et al.*, 2017). Only in stomach cancer that I observe an association between older age and the presence of MSI-H tumours, as reported previously (Cancer Genome Atlas Research, 2014a) (OR = 1.0392, 95%CI = 1.0091-1.0720, *p*-value = 0.01). When I further examined associations between age and mutations in *POLE* and *POLD1* in other cancers apart from endometrial cancer, no significant associations were observed.

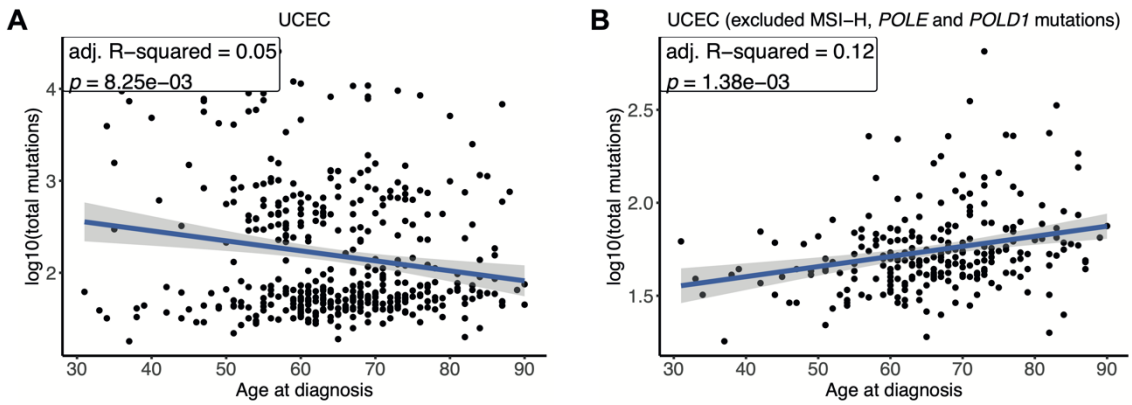


Figure 3.12. Association between patient's age and somatic mutation burden in endometrial cancer.

Association between age and mutation burden in endometrial cancer for **(A)** all samples and **(B)** samples excluding tumours with MSI-H and *POLE/POLD1* mutations. Simple linear regression was performed to investigate the association between age and mutation burden. Further investigation was carried out using multiple linear regression. Adjusted R-squared and *p*-values before multiple hypothesis testing correction from multiple linear regression analyses are displayed. Blue lines represent the best linear fit. The Grey envelope denotes a 95% confidence interval.

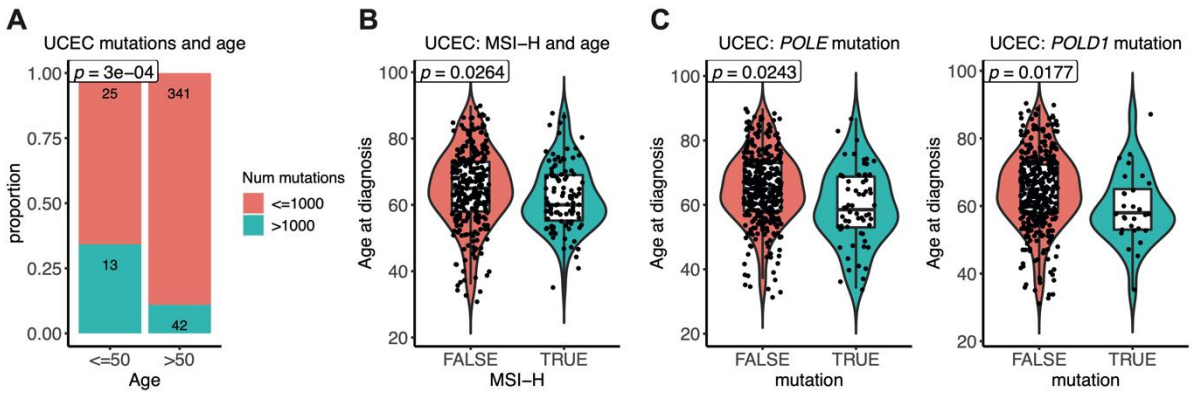


Figure 3.13. Hypermutated tumours in endometrial cancer and their association with patient's age.

(A) The proportion of hypermutated tumours (>1,000 mutations/exome) in young (age ≤ 50) and old (age > 50) endometrial cancer. The statistical significance (*p*-value) was calculated using a two-sided Fisher's exact test. **(B)** The association between age and MSI-H in endometrial cancer. The statistical significance was calculated from the multiple logistic regression adjusting for clinical variables. The *p*-value is shown in the figure. **(C)** The association between age and *POLE/POLD1* mutations in endometrial cancer. The statistical significance (*p*-value) was calculated from the multiple logistic regression adjusting for clinical variables. The middle bar of the boxplot is the median. The box represents the interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of 1.5 x IQR.

Although the increase in mutation load with age in cancer is well studied (Milholland *et al.*, 2015; Chalmers *et al.*, 2017), differences in mutation rate in particular genes as a function of age across cancer types are largely unknown. To bridge this gap of knowledge, I conducted logistic regression to investigate genes that are more or less likely to be mutated with increased age. To prevent potential biases caused by hypermutated tumours, I restricted the analysis to samples with < 1,000 non-silent mutations per exome and those that are not MSI-H tumours (**Table 3.1**). I first investigated associations between age and pan-cancer gene-level mutations. Using multiple logistic regression correcting for gender, race, and cancer type, mutations in *IDH1* (OR = 0.9608, 95%CI = 0.9497-0.9719, adj. *p*-value = 1.73 x 10^{-10}) and *ATRX* (OR = 0.9804, 95%CI = 0.9725-0.9884, adj. *p*-value = 1.60x10^{-5}) showed a negative association with age. On the other hand, mutations in *PIK3CA* were more common in older individuals (OR = 1.0096, 95%CI = 1.0035-1.0158, adj. *p*-value = 0.0139) (**Figure 3.14A**). I next identified genes exhibiting mutation rate differences associated with age in a cancer-specific manner in 24 cancer types with at least 100 samples (**Table 3.1**). Using logistic regression, I identified 31

mutations in 12 cancer types that increased or decreased as a function of the patients' age (adj. *p*-value < 0.05) (**Figure 3.14B, Table 3.3**). The most striking negative associations between mutations and age in low-grade glioma and glioblastoma were found in *IDH1* (OR = 0.9509 and 0.8962, 95%CI = 0.9328-0.9686 and 0.8598-0.9291, adj. *p*-value = 4.12×10^{-7} and 1.78×10^{-9}, respectively), *ATRX* (OR = 0.9471 and 0.9120, 95%CI = 0.9310-0.9628 and 0.8913-0.9466, adj. *p*-value = 1.67×10^{-10} and 2.33×10^{-8}, respectively), and *TP53* (OR = 0.9431 and 0.9736, 95%CI = 0.9274-0.9582 and 0.9564-0.9905, adj. *p*-value = 1.08×10^{-12} and 5.16×10^{-3}, respectively). This observation was consistent with the fact that the median age of IDH-mutants is younger than IDH-WT gliomas. Patients carrying *IDH1* mutations generally had longer survival than IDH-WT patients (Yan *et al.*, 2009). Previous studies also reported that *IDH1* mutations often co-occurred with *ATRX* and *TP53* mutations, and mutations in these three genes were more prevalent in gliomas without *EGFR* mutations (Brennan *et al.*, 2013; Cancer Genome Atlas Research *et al.*, 2015). Indeed, I found that *EGFR* mutations were more common in older low-grade glioma patients (OR = 1.0865, 95%CI = 1.0525-1.1258, adj. *p*-value = 4.13×10^{-7}) (**Figure 3.15**). Moreover, the SCNA analysis revealed an increase in focal gains of *EGFR* with age in low-grade glioma but not in glioblastoma (**Figure 3.10A**), suggesting differences in the age-associated genomic landscape between the two glioma types. Together with the SCNA results, gliomas from younger patients are associated with *IDH1*, *ATRX*, and *TP53* mutations, lower SCNAs, and longer survival. In contrast, gliomas from older patients were more likely to be IDH-WT with *EGFR* mutations, chromosome 7 gain and 10 loss, *CDKN2A* deletion and worse prognosis. This clearly highlights biological differences between age-related subtypes in gliomas.

Table 3.3: List of cancer driver genes that show age-associated mutation patterns

Cancer type	Gene	Odds ratio	95% Confidence interval	p-value	Adjusted p-value	Direction
BLCA	*ELF3*	1.0602	1.0228-1.1017	1.16×10^{-3}	3.16×10^{-3}	Increase
BRCA	*CDH1*	1.0218	1.0049-1.3921	1.11×10^{-2}	1.83×10^{-2}	Increase
BRCA	*GATA3*	0.9697	0.9526-0.9866	4.33×10^{-4}	1.37×10^{-3}	Decrease
BRCA	*KMT2C*	1.0381	1.0187-1.0583	8.86×10^{-5}	3.74×10^{-4}	Increase
BRCA	*MAP3K1*	1.0217	1.0022-1.0418	2.88×10^{-2}	3.77×10^{-2}	Increase
BRCA	*PIK3CA*	1.0128	1.0010-1.0249	3.36×10^{-2}	4.11×10^{-2}	Increase
CESC	*EP300*	1.0352	1.0069-1.0649	1.45×10^{-2}	2.30×10^{-2}	Increase
CESC	*NOTCH1*	1.0447	1.0038-1.0883	3.23×10^{-2}	4.09×10^{-2}	Increase
CESC	*PIK3CA*	1.0347	1.0129-1.0577	1.60×10^{-3}	4.05×10^{-3}	Increase
CESC	*PTEN*	1.0550	1.0174-1.0959	3.86×10^{-3}	7.33×10^{-3}	Increase
CESC	*TP53*	1.0539	1.0192-1.0921	1.98×10^{-3}	4.70×10^{-3}	Increase
GBM	*ATRX*	0.9198	0.8913-0.9466	2.45×10^{-9}	2.33×10^{-8}	Decrease
GBM	*IDH1*	0.8962	0.8598-0.9291	1.41×10^{-10}	1.78×10^{-9}	Decrease
GBM	*TP53*	0.9736	0.9564-0.9905	2.31×10^{-3}	5.16×10^{-3}	Decrease
HNSC	*CASP8*	1.0623	1.0311-1.0968	4.13×10^{-5}	1.96×10^{-4}	Increase
HNSC	*CREBBP*	1.0461	1.0054-1.0912	2.55×10^{-2}	3.59×10^{-2}	Increase
HNSC	*FAT1*	1.0399	1.0175-1.0637	3.53×10^{-4}	1.22×10^{-3}	Increase
HNSC	*FBXW7*	1.0417	1.0052-1.0819	2.43×10^{-2}	3.55×10^{-2}	Increase
HNSC	*NOTCH1*	1.0293	1.0050-1.0550	1.74×10^{-2}	2.64×10^{-2}	Increase
LGG	*ATRX*	0.9472	0.9310-0.9628	8.77×10^{-12}	1.67×10^{-10}	Decrease
LGG	*EGFR*	1.0865	1.0525-1.1258	6.52×10^{-8}	4.13×10^{-7}	Increase
LGG	*FUBP1*	1.0337	1.0095-1.0588	6.32×10^{-3}	1.09×10^{-2}	Increase
LGG	*IDH1*	0.9509	0.9328-0.9686	5.42×10^{-8}	4.12×10^{-7}	Decrease
LGG	*TP53*	0.9431	0.9274-0.9582	2.83×10^{-14}	1.08×10^{-12}	Decrease
LUAD	*FLG*	0.9588	0.9348-0.9828	8.03×10^{-4}	2.35×10^{-3}	Decrease
PCPG	*NF1*	1.0416	1.0045-1.0837	2.75×10^{-2}	3.73×10^{-2}	Increase
PRAD	*FOXA1*	1.0977	1.0268-1.1795	5.65×10^{-3}	1.02×10^{-2}	Increase
SKCM	*BRAF*	0.9706	0.9545-0.9861	1.82×10^{-4}	6.93×10^{-4}	Decrease
SKCM	*NF1*	1.0643	1.0321-1.1018	2.60×10^{-5}	1.41×10^{-4}	Increase
STAD	*CDH1*	0.9414	0.9027-0.9800	3.02×10^{-3}	6.03×10^{-3}	Decrease
UCEC	*CTNNB1*	0.9528	0.9210-0.9836	2.60×10^{-3}	5.50×10^{-3}	Decrease

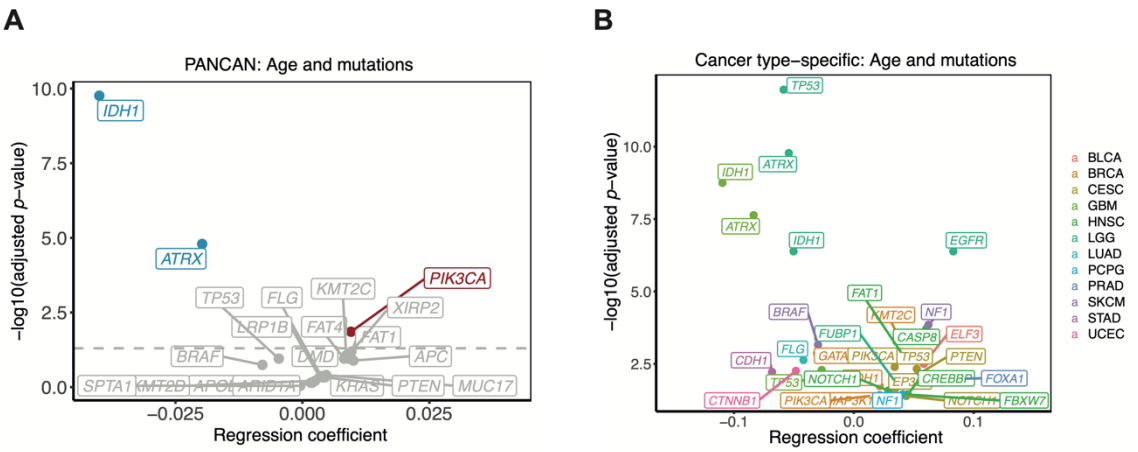


Figure 3.14. Association between cancer patients' age and somatic mutations in cancer driver genes.

(A) A pan-cancer association between age and mutations. Multiple logistic regression coefficients and significant values are shown. Genes with a significant positive and negative association between age and somatic mutations after using multiple logistic regression (adj. *p*-value < 0.05) are highlighted in red and blue, respectively. **(B)** Summary of the cancer type-specific association between age and mutations. Multiple logistic regression coefficients and significant values are shown. Only genes with a significant association (adj. p-value < 0.05) are shown in the figure. A colour code is provided to denote the cancer type where the association between age and gene mutation was found.

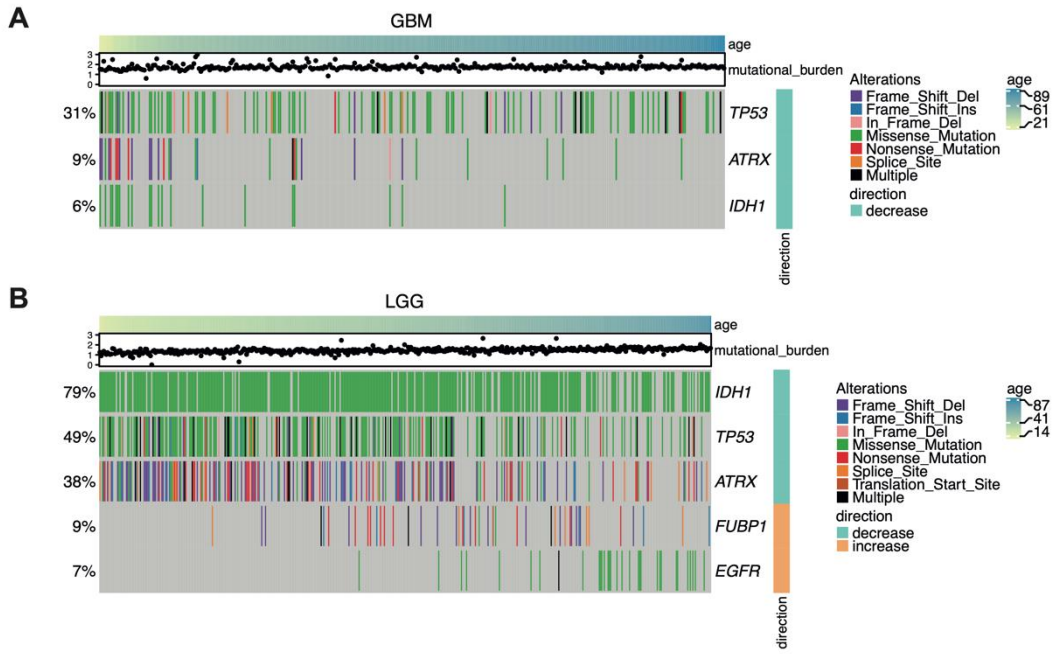


Figure 3.15. Examples of age-associated somatic mutations.

Heatmap showing age-associated mutations in **(A)** glioblastoma and **(B)** low-grade glioma. Samples are sorted by age. Colours represent types of mutation. The right annotation legend indicates the direction of change, increase or decrease mutations with age. The mutation burden of samples is presented in the dot above the heatmap on a log10 scale.

Mutations in *CDH1* were more frequent in younger stomach cancer patients (OR = 0.9414, 95%CI = 0.9027-0.9800, adj. *p*-value = 0.006) (**Figure 3.16A**), but more common in older breast cancer patients (OR = 1.0218, 95%CI = 1.0049-1.0392, adj. *p*-value = 0.0183) (**Figure 3.17A**). This result highlights cancer-specific patterns of genomic alterations with age. I tested whether age-associated subtypes could explain differences in mutation with age. Using subtype information from a previous TCGA study (Liu *et al.*, 2018b), *CDH1* mutations were found more often in the genomically stable (GS) subtype of stomach cancer (two-sided Fisher's exact, *p*-value = 2.0×10^{-5}), which was presented more frequently in younger patients (two-sided Wilcoxon rank-sum test, *p*-value = 0.0058) (**Figure 3.16B-C**). As expected, *CDH1* mutations were highly enriched in the invasive lobular carcinoma (ILC) subtype of breast cancer (Berger *et al.*, 2018) (two-sided Fisher's exact, *p*-value = 4.4×10^{-38}), which was more prevalent in older patients (two-sided Wilcoxon rank-sum test, *p*-value = 0.00081) (**Figure 3.17B-C**). Overall, these results demonstrate that non-silent mutations in cancer driver genes were not uniformly distributed across ages, and I have comprehensively identified, based on data available at present, cancer driver genes that show age-associated mutation patterns, which might partly be

explained by the presence of age-related subtypes in some cancers. These patterns might point out age-associated disparities in carcinogenesis, molecular subtypes and survival outcome.

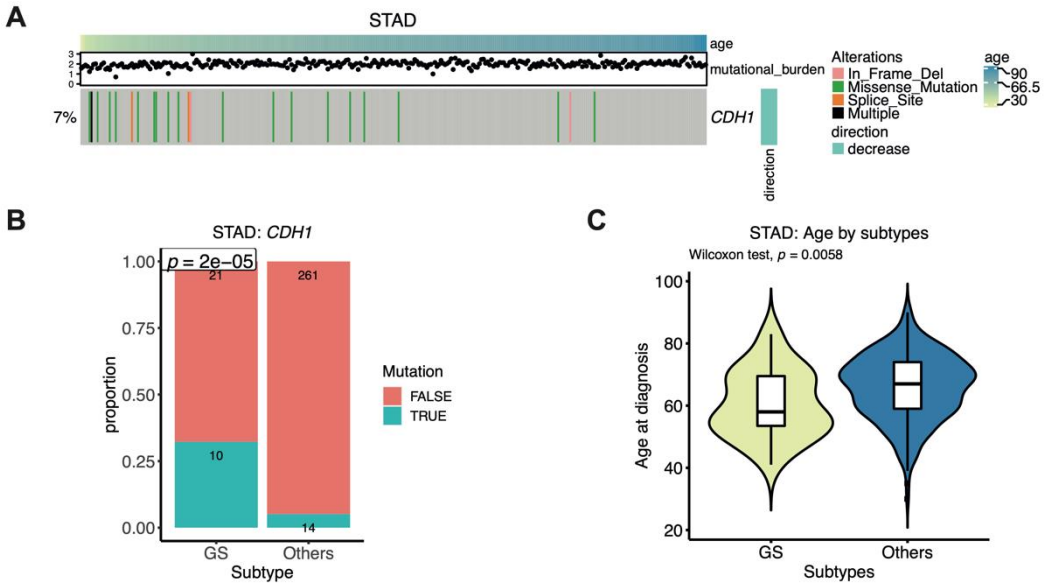


Figure 3.16. Age-related CDH1 mutation in stomach cancer.

(A) Heatmap showing age-associated mutations in stomach cancer. Samples are sorted by age. Colours represent types of mutation. The right annotation legend indicates the direction of change, increase or decrease mutations with age. The mutation burden of samples is presented in dots above the heatmap on a log10 scale. **(B)** Association between *CDH1* mutations and stomach cancer subtype comparing between genomically stable (GS) and other subtypes. The statistical significance (*p*-value shown) was calculated from the two-sided Fisher's exact test. **(C)** Age distribution of genomically stable stomach cancer subtype and other stomach cancer subtypes. GS *n* = 31, Others *n* = 275 samples. The statistical significance (*p*-value shown) was calculated from the two-sided Wilcoxon rank-sum test. The middle bar of the boxplot is the median. The box represents the interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of 1.5 x IQR.

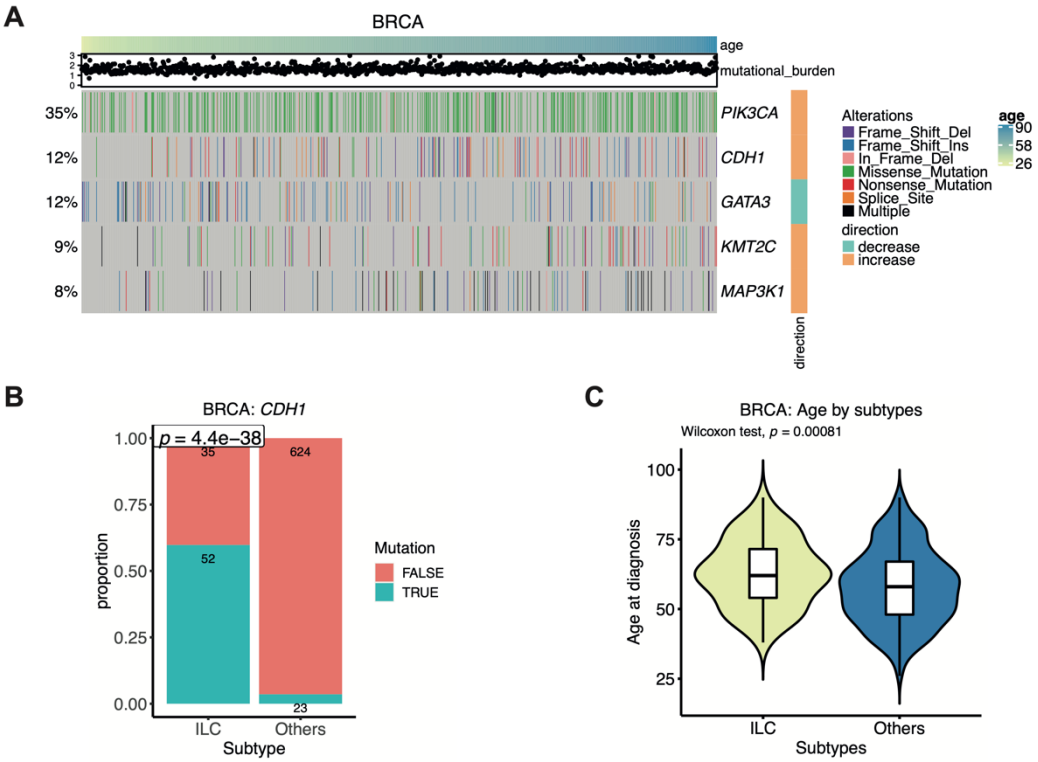


Figure 3.17. Age-related CDH1 mutation in breast cancer.

(A) Heatmap showing age-associated mutations in breast cancer. Samples are sorted by age. Colours represent types of mutation. The right annotation legend indicates the direction of change, increase or decrease mutations with age. The mutation burden of samples is presented in dots above the heatmap on a log10 scale. **(B)** Association between *CDH1* mutations and breast cancer subtype comparing between invasive lobular carcinoma (ILC) and other subtypes. The statistical significance (*p*-value shown) was calculated from the two-sided Fisher's exact test. **(C)** Age distribution of invasive lobular carcinoma breast subtype and other breast cancer subtypes. ILC *n* = 647, Others *n* = 87 samples. The statistical significance (*p*-value shown) was calculated from the two-sided Wilcoxon rank-sum test. The middle bar of the boxplot is the median. The box represents the interquartile range (IQR), 25th to 75th percentile. Whiskers represent a distance of 1.5 x IQR.

3.3.4 Age-associated alterations in oncogenic signalling pathways

As I have identified numerous age-associated alterations in cancer driver genes both at the level of somatic mutations and SCNAs, I asked if the age-associated patterns also exist in particular oncogenic signalling pathways. I used the data from a previous TCGA study, which had comprehensively characterized 10 highly altered signalling pathways in cancers (Sanchez-Vega *et al.*, 2018). To make the subsequent analysis comparable to previous analyses, I

restricted the analysis to samples that were used in the previous analyses, yielding 8,055 samples across 33 cancer types (**Table 3.1**). Member genes in the pathways were accessed for SCNAs, mutations, epigenetic silencing through promoter DNA hypermethylation, and gene fusions. Using logistic regression adjusting for gender, race, and cancer type, I identified five out of 10 signalling pathways that showed a positive association between pathway alterations and patient's age (adj. *p*-value < 0.05), indicating that the genes in these pathways are altered more frequently in older patients, concordant with the increase in overall mutations and SCNAs with age (**Figure 3.18A**). The strongest association was found in cell cycle (OR = 1.0122, 95%CI = 1.0076-1.0168, adj. *p*-value = 1.40×10^{-6}) and Wnt signalling (OR = 1.0122, 95%CI = 1.0073-1.0172, adj. *p*-value = 6.39×10^{-6}) pathways. I next applied logistic regression to investigate the cancer-specific association between age and oncogenic signalling alterations for cancer types that contained at least 100 samples. In total, I identified 28 significant associations across 15 cancer types (adj. *p*-value < 0.05) (**Figure 3.18B**). Alterations in Hippo and TP53 signalling pathways are significantly associated with age, both positively and negatively, in five cancer types. Consistent with the pan-cancer analysis, cell cycle, Notch and Wnt signalling each showed an increase in alterations with age in three cancer types. I found that alterations in cell cycle pathway increased with age in low-grade glioma (OR = 1.0313, 95%CI = 1.0161-1.0467, adj. *p*-value = 0.00035). This was largely explained by the increase in *CDKN2A* and *CDKN2B* deletions with age, as well as epigenetic silencing of *CDKN2A* in older patients (**Figure 3.19A**). On the other hand, TP53 pathway alteration was more pronounced in younger patients (OR = 0.9520, 95%CI = 0.9372-0.9670, adj. *p*-value = 2.63×10^{-8}), somatic mutations in the *TP53* gene account for this observation (**Figure 3.19B**). In endometrial cancer, two pathways – Hippo (OR = 0.9681, 95%CI = 0.9459-0.9908, adj. *p*-value = 0.0126) and Wnt (OR = 0.9741, 95%CI = 0.9541-0.9946, adj. *p*-value = 0.0240) - showed a negative association with age, that may be explained by the presence of hypermutated tumours in younger patients. Collectively, these results provide for the first time an overview of pathway alterations in relation to age in several cancer types, highlighting differences in oncogenic pathways that might be important in cancer initiation and progression in an age-related manner.

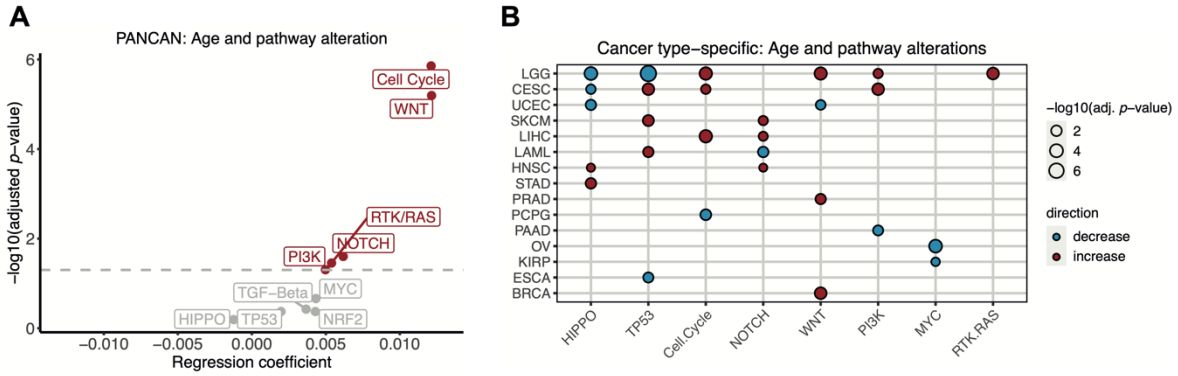


Figure 3.18. Association between cancer patients' age and oncogenic signalling pathway alterations.

(A) Association between age and oncogenic pathway alterations in the pan-cancer level. Multiple logistic regression coefficients and significant values are shown. Pathways with a significant positive association between age and alterations (adj. *p*-value < 0.05) are highlighted in red. **(B)** Cancer-specific age-associated pathway alterations. Pathways that show a significant positive and negative association with age per cancer type (multiple logistic regression, adj. *p*-value < 0.05) are displayed in red and blue dots, respectively.

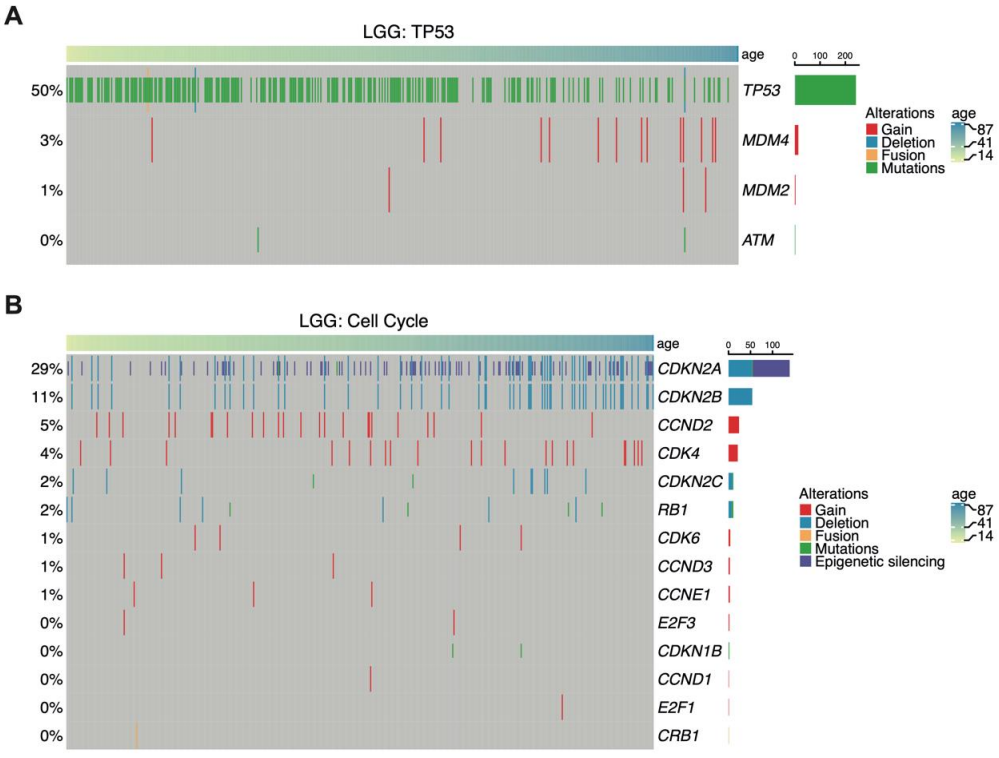


Figure 3.19. Examples of age-associated oncogenic signalling pathways.

Heatmap showing age-associated alterations in genes associated with **(A)** TP53 and **(B)** cell cycle pathways in low-grade glioma. Samples are sorted by age. Colours represent types of alteration.

3.3.5 Age-associated gene expression and DNA methylation changes

In addition to the genomic differences with age, I also investigated age-associated transcriptomic and epigenetic changes across cancers. I separately performed multiple linear regression analyses on gene expression data and methylation data of 24 cancer types that contained at least 100 samples in both types of data (**Table 3.1**). I noticed that, across all genes, the regression coefficient of age on gene expression negatively correlated with the regression coefficient of age on methylation in all cancer types (**Figure 3.20**), suggesting that global changes of gene expression and methylation with age are in the opposite direction. This supports the established role of DNA methylation in suppressing gene expression. Numbers of significant differentially expressed genes with age (age-DEGs) (adj. p-value < 0.05) varied from nearly 5,000 up- and down-regulated genes in low-grade glioma to no significant gene in 5 cancer types, including pancreatic cancer, sarcoma, stomach cancer, testicular germ cell tumour, and thyroid adenocarcinoma. By applying a similar linear regression model, I identified significant differentially methylated genes with age (age-DMGs) (adj. p-value < 0.05) across cancer types. Notably, the numbers of age-DEGs and age-DMGs were consistent for most cancer types (**Figure 3.21**). It is worth noting that cancers of female reproductive organs, including breast, ovarian and endometrial cancers, were among the highest number of age-DEGs and age-DMGs. Consistently, these female-related cancers were in the age-associated cancer group identified using survival and gene expression analyses by a recent study (Shah *et al.*, 2021).

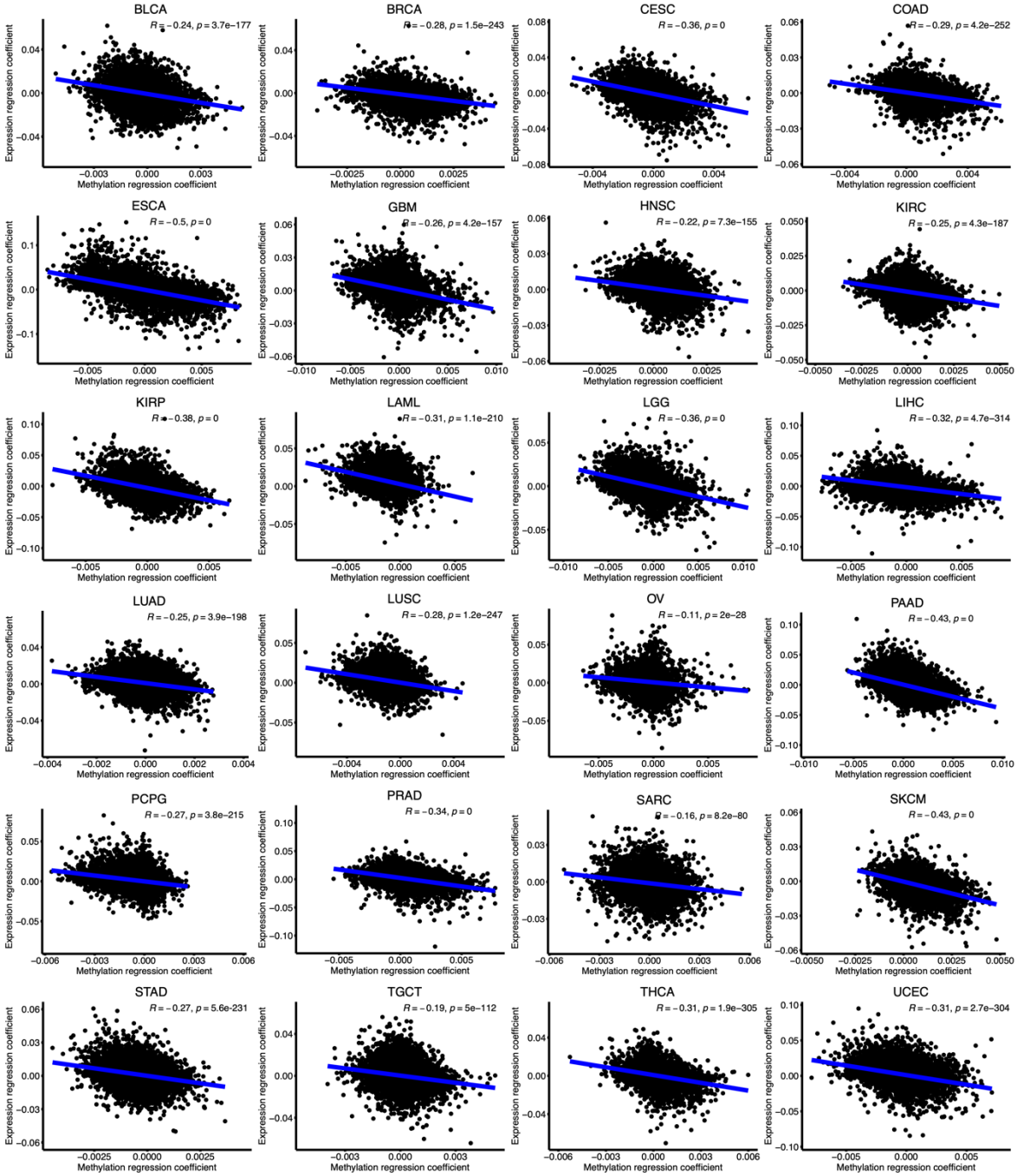


Figure 3.20. Negative association between age-related gene expression and DNA methylation.

Pearson correlation between linear regression coefficient of age on DNA methylation level and linear regression coefficient of age on gene expression. The regression coefficients were obtained from the multiple linear regression analysis to investigate the association between age and DNA methylation or gene expression. Pearson correlation coefficient r and *p*-values (two-sided test) are shown in the figures.

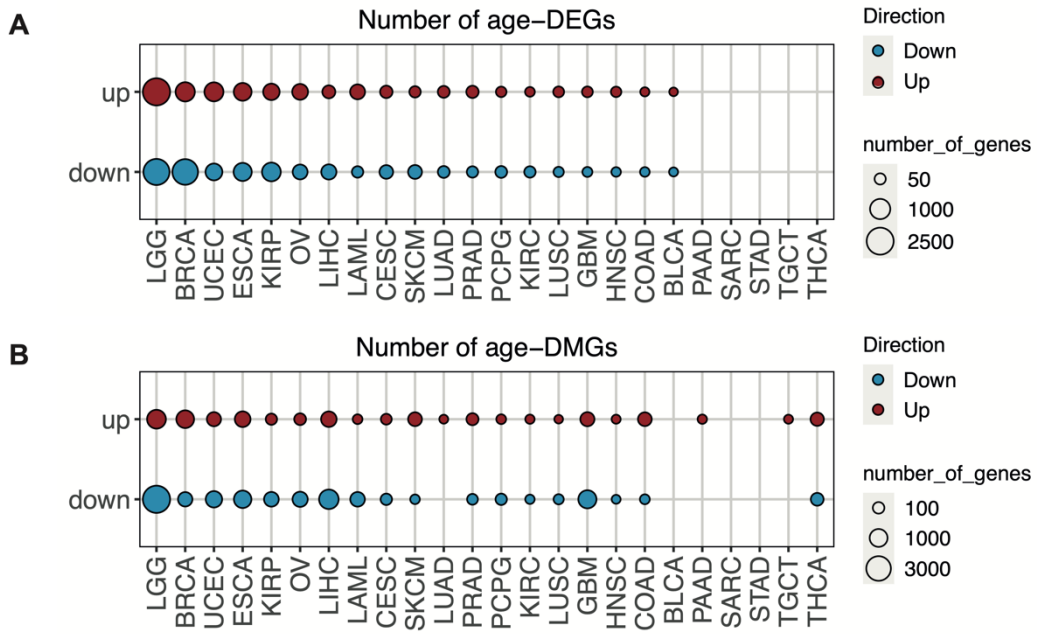


Figure 3.21. Age-related gene expression and DNA methylation changes in cancer.
The number of **(A)** age-DEGs and **(B)** age-DMGs across cancer types. Red dots represent up-regulated genes, while blue dots denote down-regulated genes. The dot size corresponds to the number of genes.

To exclude the possibility that germline predisposition mutations in some patients may cause such a high number of age-DEGs and age-DMGs in cancers of the female reproductive system, I excluded samples harbouring germline mutations in *BRCA1*, *BRCA2* and *TP53* as previously identified (Huang *et al.*, 2018) from the breast, ovarian, and endometrial cancer cohorts and repeated the multiple linear regression analysis. I observed a high correlation between regression coefficients of the analyses from all tumours and the analyses excluding samples with germline mutations for all three cancer types and both gene expression and methylation ($R = 0.93\text{-}0.99$, p-value $< 2.2 \times 10^{-16}$) (**Figure 3.22A**). The overlap between age-DEGs or age-DMGs identified from all samples and from samples without germline variants were large (**Figure 3.22B**). Therefore, the high number of age-DEGs and age-DMGs are independent of the presence of germline predisposition mutations in some patients. I, thus, used age-DEGs and age-DMGs identified from all samples for subsequent analyses.

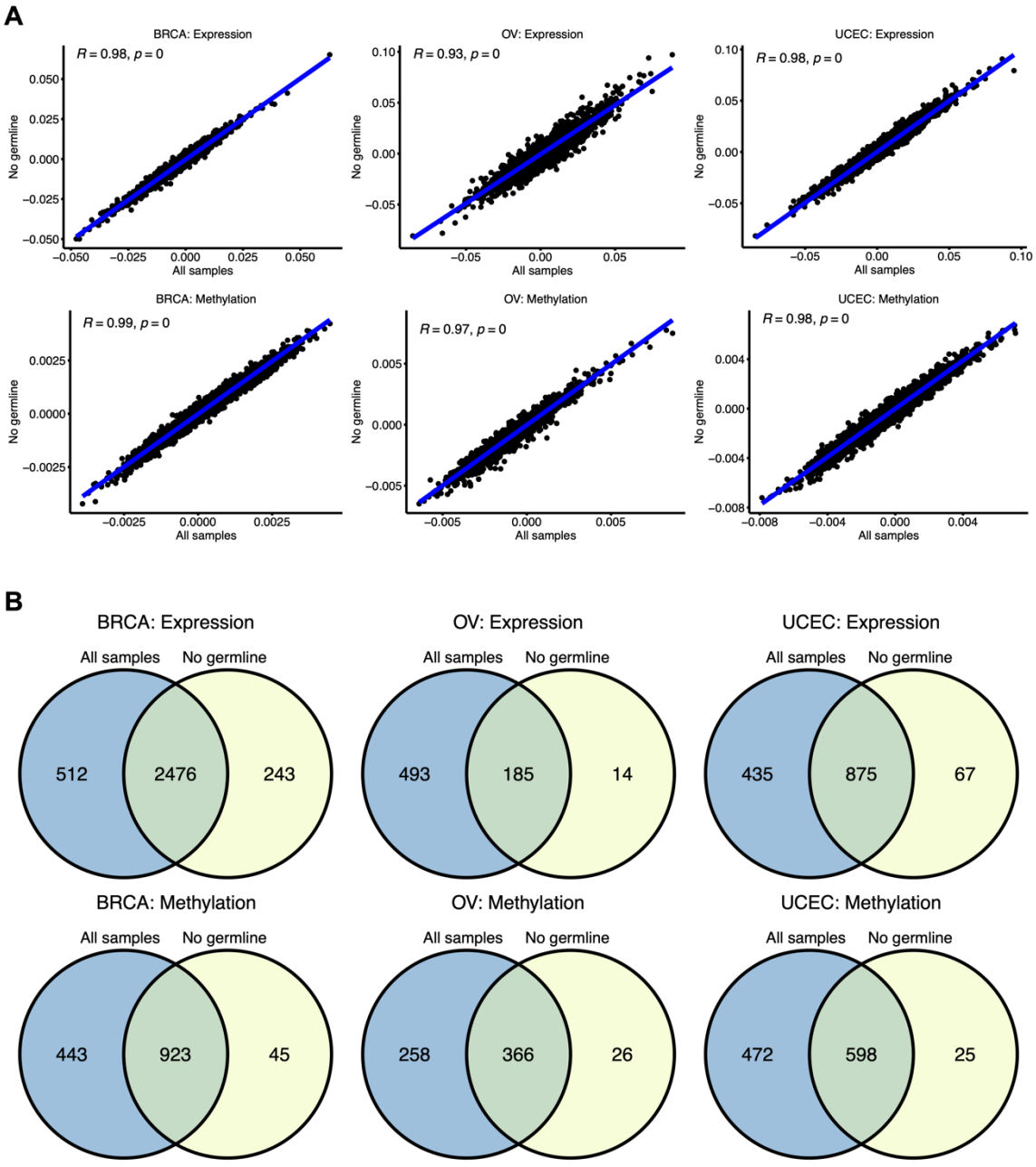


Figure 3.22. Impact of germline predisposition mutations on age-associated gene expression and DNA methylation in breast, ovarian, and endometrial cancers.

(A) Pearson correlation between linear regression coefficient of age on gene expression or methylation from all samples and only samples without germline predisposition mutations. **(B)** Overlap between age-DEGs or age-DMGs identified from all samples and only samples without germline predisposition mutations.

I next focused subsequent analyses on 10 cancer types that contained at least 150 age-DEGs and 150 age-DMGs, including low-grade glioma, breast cancer, endometrial cancer, oesophageal cancer, papillary renal cell carcinoma, ovarian cancer, liver cancer, acute myeloid leukaemia, melanoma, and prostate cancer. I identified overlapping genes between age-DEGs and age-DMGs and found that most of them, from 84% (37/44 genes) in ovarian cancer to 100% in acute myeloid leukaemia (57 genes) and prostate cancer (7 genes), were genes that presented increased methylation and decreased expression with age and genes that had decreased methylation and increased expression with age (**Figure 3.23**). I further examined the correlation coefficient between methylation and expression comparing between 4 groups of genes 1) genes overlapping between age-DMGs and age-DEGs (age-DMGs-DEGs), 2) age-DMGs only, 3) age-DEGs only, and 4) other genes. I found that age-DMGs-DEGs had the most negative correlation between DNA methylation and expression when compared with other groups of genes (**Figure 3.24**), highlighting that age-associated gene expression changes in cancer are repressed, at least in part, by DNA methylation.

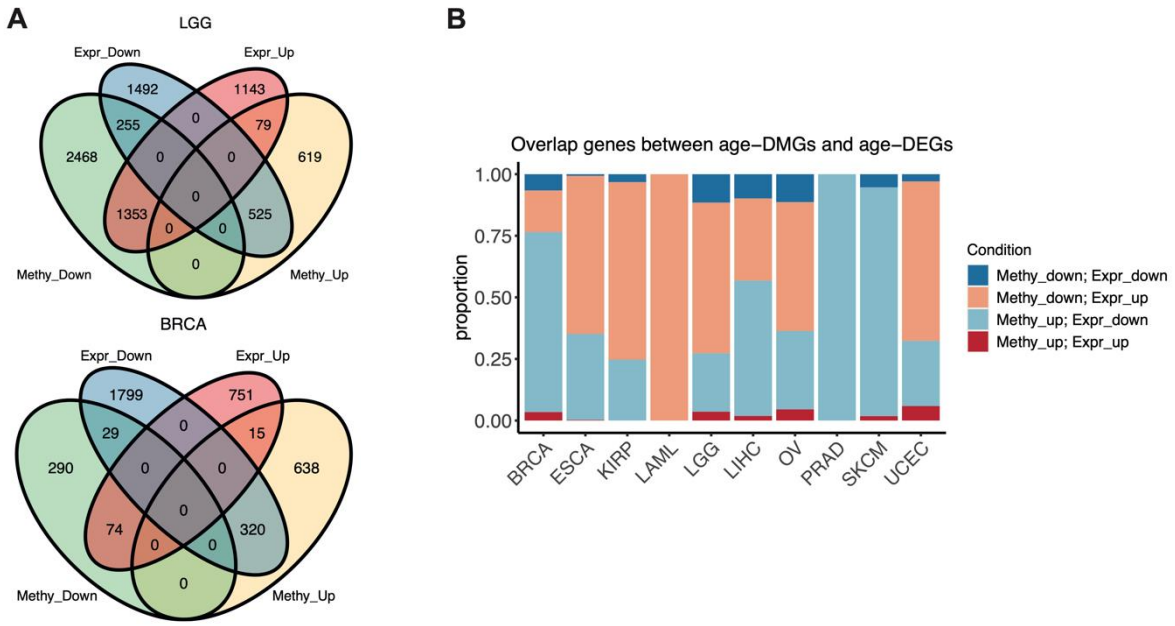


Figure 3.23. Opposite direction of age-related gene expression and DNA methylation changes in cancer.

(A) Venn diagrams of the overlap between age-DEGs and age-DMGs in low-grade glioma and breast cancer are shown as examples. **(B)** The distribution of overlap genes between age-DMGs and age-DEGs. The genes were classified into (1) down-regulated methylation and down-regulated expression, (2) down-regulated methylation and up-regulated expression, (3) up-regulated methylation and down-regulated expression, and (4) up-regulated methylation and up-regulated expression.

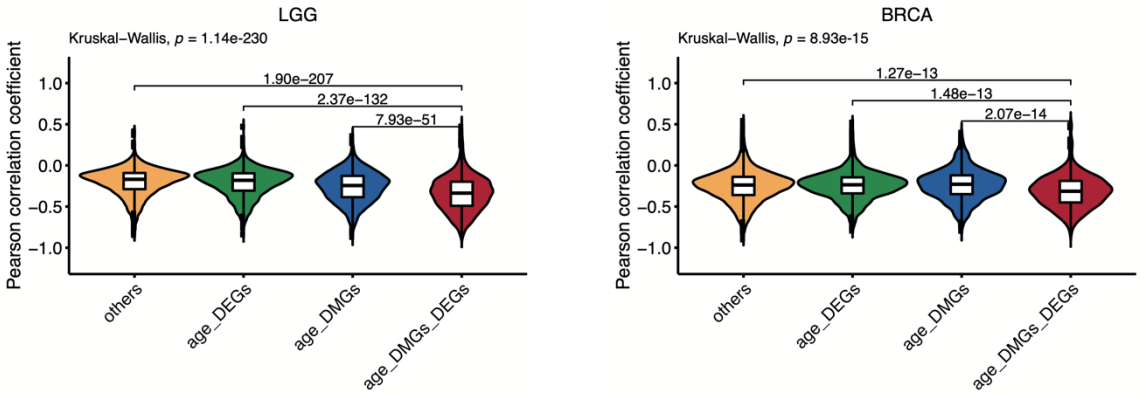


Figure 3.24. Age-related gene expression in cancers was partly controlled by age-related DNA methylation.

Violin plots showing the distribution of the Pearson correlation coefficient between methylation and gene expression in low-grade glioma and breast cancer. Genes were grouped into (1) common genes between age-DMGs and age-DEGs (age-DMGs-DEGs), (2) age-DMGs only genes, (3) age-DEGs only genes, and (4) other genes. The group comparison was performed by the Kruskal-Wallis test. The pairwise comparisons were done using the two-sided Dunn's test. *P*-values from Dunn's test between age-DMGs-DEGs and the other groups adjusted by Bonferroni correction are shown.

I next performed Gene Set Enrichment Analysis (GSEA) to gain biological insights into the expression and methylation changes with age. I identified various significantly enriched Gene Ontology (GO) terms across cancers (adj. *p*-value < 0.05 for gene expression and adj. *p*-value < 0.1 for methylation) (**Figure 3.25**). Notably, several GO terms were enriched in both expression and methylation changes in the opposite direction. The enriched terms in breast cancer included several signalling, metabolism, and developmental pathways. The Wnt signalling pathway, which was altered more frequently in older breast cancer patients (**Figure 3.18B**), showed decreased gene expression and increased methylation with age. Interestingly, in low-grade glioma, mitochondrial terms were enriched in the gene expression of older patients. Mitochondrial dysfunction is known to be important in glioma pathophysiology (Ordys *et al.*, 2010); thus, the different levels of mitochondrial aberrations might contribute to disparities in the aggressiveness of gliomas in patients of different ages. I also identified numerous immune-related terms enriched across several cancer types, including oesophageal, papillary renal cell, liver, and prostate cancers (**Figure 3.25**). Recent studies suggested alterations in immune-related gene expression and immune cell abundance changes with age in cancers (Erbe *et al.*, 2021; Shah *et al.*, 2021; Wu *et al.*, 2021). For instance, a recent study

reported decreased T cell and increased macrophage abundance in older breast cancer patients (Erbe *et al.*, 2021). Understanding the immune-landscape differences between younger and older patients might be beneficial for selecting treatment strategies for immunotherapy in the future.

Taken together, I have systematically characterised the transcriptome and methylation in relation to age across cancer types. These results suggest that gene expression changes with age in cancer are controlled, at least in part, by DNA methylation. These changes reflect differences in biological pathways that might be important in tumour development.

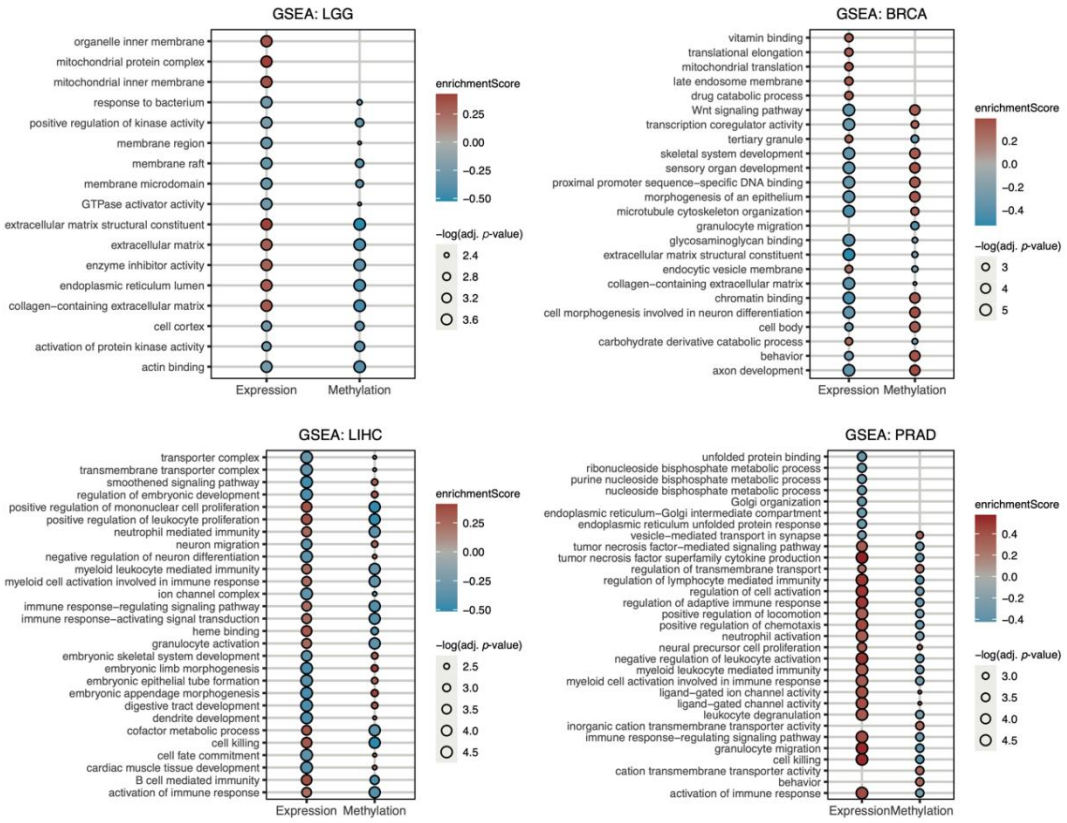


Figure 3.25. Gene Set Enrichment Analysis (GSEA) of age-related gene expression and DNA methylation.

The enriched gene ontology (GO) terms identified by GSEA in low-grade glioma, breast cancer, liver cancer, and prostate cancer. The dot size corresponds to a significant level. A GO term was considered significantly enriched term if adj. *p*-value < 0.05 for gene expression and adj. *p*-value < 0.1 for methylation. Colours represent enrichment scores, red denotes positive scores (enriched in older patients), and blue signifies negative scores (enriched in younger patients).

3.4 Discussion

Although age is an important risk factor for cancer, how age impacts the molecular landscape of cancer is not well understood. This chapter provides a comprehensive multi-omics overview of the age-associated molecular landscape in cancer, including genomic instability, LOH, WGD, SCNAs, somatic mutations, pathway alterations, gene expression and DNA methylation. I confirmed the known increase in mutation load (Alexandrov *et al.*, 2015; Milholland *et al.*, 2015), which can be in part explained by an increase in C>T mutations and found an increase in genomic instability, LOH and WGD with age in several cancer types. I identified several age-related pan-cancer and cancer-specific alterations. The highest age-related differences were evident in low-grade glioma and endometrial cancer, which could be attributed to age-related subtypes in cancer.

3.4.1 Age-related genomic alterations across cancers

Cancer develops through the accumulation of genetic and epigenetic alterations. Mutation accumulation with age is thought to be a cause of cancer, and a substantial portion of mutations arise before cancer initiation (Tomasetti *et al.*, 2013). The age-associated mutation accumulation has been demonstrated in both cancer (Alexandrov *et al.*, 2015; Milholland *et al.*, 2015) and normal tissues (Xie *et al.*, 2014; Martincorena *et al.*, 2015; Martincorena *et al.*, 2018), providing a better understanding of early carcinogenic events. In this chapter, I confirmed an increase in somatic mutations with age in the pan-cancer analysis and in most cancer types. This observation could represent the accumulation of somatic mutation with age in tissues that occur before carcinogenesis, as recently reported widely in most non-cancerous human tissues (Yizhak *et al.*, 2019; Kakiuchi & Ogawa, 2021; Moore *et al.*, 2021). In addition to somatic mutations, SCNAs, LOH and WGD increase with age in several cancers, particularly low-grade glioma, endometrial and ovarian cancers. Interestingly, SCNAs are rarely detected in non-cancerous tissues (Brunner *et al.*, 2019; Yokoyama *et al.*, 2019; Li *et al.*, 2021a; Moore *et al.*, 2021). On the contrary, SCNAs are widespread among cancer (Beroukhim *et al.*, 2010). Thus, SCNAs could be an important feature distinguishing cancer from normal cells. Recent evidence suggests that SCNA burden, not somatic mutation level, is a prognostic factor associated with recurrence and death (Hieronymus *et al.*, 2018; Smith & Sheltzer, 2018). Thus, an increased SCNA level with age could partly account for a poorer prognosis in the elderly.

When considering the cancer-specific age-related somatic mutations, there was a strong negative association between age and somatic mutation in *IDH1*, *ATRX* and *TP53* in glioma. This observation points towards the difference of patient age at diagnosis between the IDH-mutant and *IDH*-WT subtypes. IDH-mutant tumours are observed in the majority of low-grade glioma and show a favourable prognosis. *IDH*-WT low-grade gliomas, on the other hand, more resemble glioblastomas and have poorer survival. Although IDH-mutant cases are a minority of tumours in glioblastoma, they are also associated with younger age (Mirchia & Richardson, 2020). A recent functional study in neural stem cells (NSCs) showed that the combination of *IDH1*, *ATRX* and *TP53* alterations blocks NSC differentiation by causing hypermethylation of CTCF motifs flanking the *SOX2* locus, disrupting chromatin looping and dysregulation of *SOX2*, an important transcription factor in self-renewal and differentiation of NSCs (Modrek *et al.*, 2017). Impaired differentiation, growth arrest evasion by mutations in *TP53*, and alternative lengthening of telomeres by *ATRX* inactivation thus cooperatively promote gliomas in younger patients. Results in this chapter, together with others (Yan *et al.*, 2009; Verhaak *et al.*, 2010), indicates that glioma shows unique age-associated subtypes. However, more research is needed to understand how age influences the evolution of glioma subtypes.

The results from this chapter also highlighted substantial age-associated differences in the endometrial cancer genome. Younger endometrial tumours associate with *POLE* and MSI-H subtypes, leading to an enrichment of hypermutated tumours. In contrast, tumours from older patients tend to harbour more SCNAs and lower mutation load. Previous studies have classified endometrial cancer into four subtypes: *POLE*, MSI-H, copy-number low and copy-number high subtypes. The *POLE* subtype and MSI-H subtype are dominated by *POLE* and defective mismatch repair mutational signatures, respectively (Cancer Genome Atlas Research *et al.*, 2013a). Conversely, the copy-number low and copy-number high subtypes had a dominant ageing-related mutational signature (Ashley *et al.*, 2019). The *POLE* and MSI-H subtypes have a favourable prognosis, while the copy-number high subtype is associated with poor survival. Therefore, endometrial cancer from younger patients is associated with *POLE* mutations, mismatch repair defects, high mutation load and better survival outcomes. However, older endometrial cancer is related to extensive SCNAs and a worse prognosis. Notably, apart from low-grade glioma and endometrial cancer, I demonstrate that other cancer types also present an age-associated genomic landscape in cancer driver genes and oncogenic signalling pathways.

Indeed, some of these age-related differences might be explained by age-related subtypes, such as the high prevalence of *CDH1* mutation in invasive lobular breast carcinoma and genomically stable stomach cancer, that are presented more often in older and younger patients, respectively. A further detailed investigation in each cancer type is required to fully distinguish age-related subtype effects from age-associated effects or even to identify a new age-related subtype. My perspective is that whether age-associated genomic differences come from age-related subtypes or age-related features, they reflect biological differences between cancers at different ages, highlighting the impact of age on the molecular profile of cancer. One limitation, however, of the age-related genomic analyses in this chapter is that I did not include information on quantitative differences of genomic alterations, such as homozygous/heterozygous loss and clonal/subclonal mutations, in the statistical models to not over-complicate the analyses. Furthermore, how and why age-associated genomic differences and age-related subtypes occur remain to be studied.

3.4.2 How ageing processes contribute to the age-related differences in cancer?

Having identified age-related differences in the molecular landscapes of various cancers, the obvious question is: what drives these differences? Indeed, an increase in somatic mutations with normal ageing could account for an age-related increase in mutation load, as mentioned above. Furthermore, accumulating evidence has underscored the importance of tissue environment changes with ageing in cancer initiation and progression (Chatsirisupachai *et al.*, 2019; Rozhok & DeGregori, 2019; Erbe *et al.*, 2020; Fane & Weeraratna, 2020). I reason that tissue environment changes during ageing might provide different selective advantages for tumours harbouring distinct molecular alterations, in turn directing tumours to different evolutionary routes. Therefore, cancers with different genomic alterations might thrive better in younger or older patients. This hypothesis is, however, waiting for experimental evidence.

Gene expression and epigenetic changes related to ageing have been studied and linked to cancer (Johnson *et al.*, 2012; Perez *et al.*, 2018; Chatsirisupachai *et al.*, 2019; Wu *et al.*, 2019). Here, I identified numerous age-associated differentially expressed genes and corresponding DNA methylation changes in a broad range of cancers. Indeed, age-DMGs-DEGs are those with the strongest negative correlation between methylation and expression when comparing with other groups, indicating that differentially expressed genes with age in cancer are partly

regulated by methylation. Gene expression and methylation changes with age link to several biological processes, showing that tumours from patients of different ages present different phenotypes. Notably, several processes identified by GSEA analysis overlap with changes in the tissue microenvironment during ageing, such as extracellular matrix reorganisation, inflammation, and developmental pathways (Fane & Weeraratna, 2020). It is unclear, however, that how the ageing tissue environment interacts with tumours and contribute to age-associated transcriptomic and epigenetic differences in cancer.

I also noticed that cancer in female reproductive organs, including breast, ovarian and endometrial cancers are among those with the highest number of age-DEGs and age-DMGs. These cancers tend to have a higher mass-normalised cancer incidence, which may reflect evolutionary trade-offs involving selective pressures related to reproduction (Silva *et al.*, 2011). Age-associated hormonal changes could also be responsible for these age-related expression differences in cancer, as evidenced by studies in breast cancer (Benz, 2008; Osako *et al.*, 2020). For instance, a recent study identified age-correlated gene expression in breast cancer and suggested that these genes are potentially regulated by oestrogen signalling, which is decreased markedly after menopause (Osako *et al.*, 2020). Further studies are required to better understand the effect of the alterations in host circulating factors on cancer.

A limitation of the gene expression and DNA methylation analyses is that I chose only one methylation probe per gene to create a one-to-one mapping between genes and probes. Other probes might also have an impact on gene expression as well as might cause noise in the GSEA analysis from methylation data. Next, although I have already included tumour purity in the linear model, it is not possible to account for the different tumour-constituent cell proportions and thus entirely exclude the influence of gene expression in non-cancerous cells such as infiltrating immune cells (Erbe *et al.*, 2021). In relation to this issue, gene expression and DNA methylation used in the present work are from bulk assays, thus cannot capture age-related differences in a single-cell resolution. In other words, I cannot distinguish whether the differences found between younger and older tumours result from an age-related shift in cell population or age-related differences in gene expression and DNA methylation within cell types. Recent single-cell atlas in murine ageing revealed cellular changes during ageing in several tissues (Tabula Muris, 2020). Another recent study also reported age-related alterations in mammary epithelia and stroma in mice, which could potentially be related to pro-tumourigenic microenvironmental changes (Li *et al.*, 2020a). However, the single-cell study

focusing on the differences between tumours in relation to age in humans is currently lacking. Further studies are required to provide a mechanistic understanding of the impact of an ageing microenvironment in shaping tumour evolution.

3.4.3 The complementary studies on pan-cancer analysis of the age-associated differences in cancer

Interestingly, several pan-cancer studies on the age-associated differences in cancer have been published during the past year (Erbe *et al.*, 2021; Shah *et al.*, 2021; Li *et al.*, 2022). First, Li *et al*. used TCGA, American Association for Cancer Research: Genomics Evidence Neoplasia Information Exchange (AACR GENIE), and the recent Pan-Cancer Analysis of Whole Genomes (PCAWG) data to study age-associated genomic differences in cancer (Li *et al.*, 2022). Results from this study are consistent with results in this chapter in several respects. Firstly, Li *et al.* also reported an increase in somatic mutations and SCNAs as a function of age. In addition, despite using slightly different statistical models and cutoffs, several similar age-associated genomic features are identified, for example, the higher frequency of *IDH1* and *ATRX* mutations in younger glioma patients. Next, Li *et al.* explored mutational timing and signatures, which suggested possible underlying mechanisms for age-associated genomic differences. They reported that tumours from older individuals tend to accumulate more clonal mutations (i.e., mutations that arise earlier in tumour evolution). Furthermore, DNA damage repair signatures were more likely to be found in older individuals. The mutational signature related to APOBEC cytidine deaminase activity also increases with age in melanoma (Li *et al.*, 2022).

In another study, Shah *et al.* used gene expression data and survival analysis to classify cancer types into age-associated (AA) and non-age-associated (NAA) cancers (Shah *et al.*, 2021). The AA cancer types included low-grade glioma, lung squamous cell carcinoma, thyroid adenocarcinoma, and cancers of female reproductive organs (breast, ovarian, endometrial cancers). They found that younger patients from AA, but not NAA, cancers are associated with accelerated ageing and senescence measured by an epigenetic clock. However, the molecular mechanisms behind this observation and why such features are limited to only AA cancers are unclear. Consistent with the result in this chapter, Shah *et al.* also found the opposite association between DNA methylation and gene expression. Finally, they also identified several age-associated somatic mutations in cancer driver genes, including several genes

identified in this chapter, such as *IDH1*, *ATRX*, and *TP53* mutations in younger low-grade glioma patients. One significant difference between this study and the results in this chapter is that Shah *et al.* reported a higher frequency of somatic mutations in numerous cancer driver genes in younger endometrial cancer patients. This, however, could be explained by the fact that younger endometrial cancer patients are enriched in hypermutation subtypes (MSI-H and *POLE/POLD1* mutations) discussed above. In this chapter, I have excluded these hypermutated tumours from the analysis of age-related somatic mutations.

Additionally, Erbe *et al.* investigated immune checkpoint blockade (ICB) biomarkers in relation to age across cancer types (Erbe *et al.*, 2021). These ICB biomarkers are primary clinical biomarkers for an increased benefit of ICB therapy. First, consistent with results from this chapter, they reported an increased mutation burden with age in most cancer types except lung adenocarcinoma and endometrial cancer. Tumours with a higher mutation load increase the chance for displaying antigens and, thus, is more likely to be recognised by T cells. In addition, they found increased expression and decreased promoter methylation of immune checkpoint genes with age. Therefore, older patients are more likely to benefit from immunotherapy based on these biomarkers. On the other hand, Erbe *et al.* also reported decreased T cell receptors and T cell abundance, together with increased macrophage abundance with age. These effects, however, are associated with reduced ICB response. Overall, this study focused on the immunological aspects of age-related differences in cancer, which will have an implication in cancer immunotherapy.

Taken together, these independent yet complementary studies, including this chapter, thus serve as a foundation for understanding age-related differences and effects on the molecular landscape of cancer and emphasise the importance of age in cancer genomic research that is particularly valuable in clinical practice.

Chapter 4

The relationship between gene expression and epigenetic changes in mammalian development and ageing

4.1 Introduction

While development and ageing seem like different and unrelated processes, they can be highly intertwined. The hyperfunction theory of ageing suggests that ageing might, in part, originate from developmental programmes (de Magalhaes, 2012; Gems, 2022). For instance, age-related phenotypes such as presbyopia, osteoporosis, and the continuous growth of the prostate gland can be viewed as the continuum of developmental programmes that could be a result of the reduced force of natural selection during ageing (de Magalhaes, 2012; Gems, 2022). In other words, some facets of ageing may arise as an unintended outcome of development. Recent studies further support the relationship between development and ageing, indicating that short rapamycin treatment in early life could increase lifespan in a range of model organisms, including mice and fruit flies (Aiello *et al.*, 2022; Shindyapina *et al.*, 2022).

In addition, several lines of evidence support the relationship between development and ageing at a molecular level. For example, changes in gene expression during ageing might exhibit a reversal trend of those changes during development, which could be related to the loss of cellular identity during ageing (Somel *et al.*, 2010; Donertas *et al.*, 2017). More recently, Izgi *et al.* analysed mouse and human gene expression datasets to show that mammalian tissue transcriptomes diverge during postnatal development and converge towards similar levels during ageing (Izgi *et al.*, 2022). The process was termed "Divergence followed by Convergence (DiCo)". They further found that tissue-specific genes vastly contributed to this DiCo pattern, highlighting that establishing tissue identity during development and ageing is associated with the loss of identity (Izgi *et al.*, 2022). However, it is unclear whether this trend starts from prenatal development. Furthermore, the contribution of epigenetics in the regulation of gene expression changes during development and ageing is largely unexplored.

This chapter aimed to investigate the relationship between gene expression and epigenetic changes during development and ageing across tissues. Here, I retrieved RNA-seq gene

expression data and histone modification data, including ChIP-seq of trimethylation of histone 3 at lysine 4 (H3K4me3) and acetylation of histone 3 at lysine 27 (H3K27ac), from the ENCODE project spanning four tissues (forebrain, hindbrain, heart, and liver) and three prenatal developmental stages in mouse, including post-conception embryonic day 11.5 (E11.5), post-conception embryonic day 15.5 (E15.5), and neonate (P0) (Consortium *et al.*, 2020; Gorkin *et al.*, 2020). For ageing data, I obtained RNA-seq and ChIP-seq data for H3K4me3 and H3K27ac of four tissues (olfactory bulb, cerebellum, heart, and liver) across three age groups (3 months, 12 months, and 29 months) from another publicly available dataset (Benayoun *et al.*, 2019). The two histone modification marks H3K4me3 and H3K27ac are both associated with "active chromatin". Specifically, H3K4me3 is associated with active promoters, while the H3K27ac mark is highly enriched in active enhancer regions. Using these datasets, I examined the links between prenatal development and ageing across tissues in mice at the transcriptomic and epigenetic levels in this chapter.

4.2 Materials and Methods

4.2.1 Data acquisition

Raw fastq RNA-seq and ChIP-seq files (H3K4me3, H3K27ac, and input controls) for mouse prenatal development and ageing were retrieved from the ENCODE project (https://www.encodeproject.org/) using *ENCODExplorer* (version 2.12.1) (Charles Joly Beauparlant, 2020) and NCBI BioProject database (accession number PRJNA281127) (Benayoun *et al.*, 2019) using *SRAtoolkit* (version 2.10.0) (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software), respectively. Each tissue and developmental stage or age group consists of 2 biological replicates of H3K4me3 and H3K27ac ChIP-seq files and 2-3 biological replicates of RNA-seq files.

4.2.2 ChIP-seq data preprocessing and analysis

The quality of raw fastq reads was accessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw read counts were trimmed using Trim Galore! (version 0.6.4) to retain only reads with Phred score > 20 and length > 25 bp (for ENCODE data) or length > 36 bp (for PRJNA281127 data). Reads were aligned to the mm10 mouse genome assembly using bowtie (version 1.1.2). The PCR duplicated reads were removed using Picard (version 2.21.2) (https://broadinstitute.github.io/picard/). SAM and BAM files were processed using *samtools* (version 0.1.19). Reads located on the ENCODE blacklist regions (Amemiya *et al.*, 2019) were filtered out. Next, the quality of ChIP-seq experiments was accessed by the cross-correlation analysis using phantompeakqualtools (version 1.2.2) (Landt *et al.*, 2012). Peak calling was performed using MACS2 (version 2.2.7) (Zhang *et al.*, 2008). For each tissue and developmental stage/age group, ChIP-seq peaks were initially called with a relaxed *p*-value threshold at $p < 0.001$ for each biological replicate, for the pooled replicate, and for pooled pseudoreplicates of true replicates (each pseudoreplicate consists of half of the reads randomly distributed from the pooled replicate) as described in the ENCODE pipeline (Gorkin *et al.*, 2020). To keep only reliable peaks, only peaks from the pooled replicate set that at least half of their length (in bases) overlapped peaks from both true biological replicates were retained. In addition, peaks from pooled replicates that at least half of their length overlapped peaks from both pooled pseudoreplicates were kept (Gorkin *et al.*, 2020).

To generate a unified peak set for each tissue (tissue-specific peak set), peaks from each tissue at every developmental stage/age group that were no more than 1,000 bp apart were merged using *bedtools* (version 2.29.2) (Quinlan & Hall, 2010). Similarly, peaks from all tissues and every developmental state/age group were merged to obtain a peak set for the cross-tissue comparison analysis. Peaks were annotated to nearby genes using ChIPseeker (version 1.26.2) (Yu *et al.*, 2015).

To compare ChIP-seq signals between samples, ChIP signals from each sample were normalised by input control from the corresponding sample by Reads Per Kilobase of transcript per Million mapped reads (RPKM) method using the *bamCompare* function from *deepTools2* (version 3.4.1) (Ramirez *et al.*, 2016). The ChIP-seq peak score for each region is thus represented by log2(RPKM) of ChIP/Input. This step results in bigwig files. Next, correlation analysis and PCA were performed on bigwig files from all samples (global analysis with the unified peak set from all samples) and samples from each tissue (tissue-specific analysis with tissue-specific peak set) using the *multiBigwigSummary* function from *deepTools2 (Ramirez et al., 2016).*

4.2.3 RNA-seq data preprocessing and analysis

The quality of raw fastq reads was accessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw read counts were trimmed using Trim Galore! (version 0.6.4) (https://github.com/FelixKrueger/TrimGalore) to retain only reads with Phred score > 20 and length > 75 bp. Reads were aligned to the mm10 mouse genome assembly using *STAR* (version 2.7.6a) (Dobin *et al.*, 2013). SAM and BAM files were processed using *samtools* (version 0.1.19) (Li *et al.*, 2009). Quantification of gene expression was performed using RSEM (version 1.3.3) (Li & Dewey, 2011). Genome annotation was obtained from GENCODE (version M23) (https://www.gencodegenes.org/mouse/release_M23.html). To compare global gene expression between samples, only genes with CPM > 1 in at least 50% of samples were retained. Reads were normalised by the trimmed mean of M values (TMM) method from the *edgeR* package (version 3.32.1) (Robinson *et al.*, 2010). PCA was performed on normalised gene expression by using *factoextra* package (version 1.0.7) (Kassambara, 2020b).

Differential expression was performed separately for development (ENCODE samples) and ageing (PRJNA281127 samples). Gene expression read counts were filtered to keep only protein-coding genes. For each tissue, only genes with a CPM > 1 in at least 50% of samples were retained. Reads were normalised by the TMM method from the *edgeR* package (version 3.32.1) (Robinson *et al.*, 2010). Differential expression analysis was performed on each pair of developmental stages (E11.5 vs E15.5, E15.5 vs P0, and E11.5 vs P0) and age groups (3M vs 12M, 12M vs 29M, and 3M vs 29M) using *edgeR* package (version 3.32.1) (Robinson *et al.*, 2010). For the comparisons between developmental stages, genes were considered differentially expressed if adjusted *p*-values < 0.05 and |Fold change| > 2. For the comparisons between age groups, genes were considered differentially expressed if adjusted *p*-values < 0.05 and |Fold change| > 1.5.

4.2.4 Analysis of tissue-specific genes

I performed differential expression analysis at each developmental stage and age group to retrieve tissue-specific genes. Samples from two brain regions, the forebrain (olfactory bulb) and hindbrain (cerebellum), were combined to analyse brain samples, resulting in three tissues, including the brain, heart, and liver. To identify tissue-specific genes, differential expression analysis comparing samples from the tissue of interest and the remaining samples was performed using the *edgeR* (Robinson *et al.*, 2010). For example, brain-specific genes were derived from differential expression analysis of brain samples compared with heart and liver samples. Genes were considered tissue-specific if adjusted p-values < 0.01 and |Fold change| > 5. Next, confident tissue-specific genes for each tissue were obtained by selecting only genes which were tissue-specific in at least two developmental stages and age groups.

4.2.5 Overlap analysis

The overlap analyses were performed using a one-sided Fisher's exact test. In each experiment, the background was set differently depending on where the genes came from. For the overlap between differentially expressed genes during development and those changes during ageing, all genes that were passed the filtering step during differential expression analysis (CPM > 1 in at least 50% of samples from that tissue) were used as a background. The overlap was considered significant if an adjusted *p*-value was < 0.05.

4.2.6 Functional enrichment analysis

Gene ontology (GO) enrichment analysis was performed using the *clusterProfiler* package in R (version 3.8.1) (Yu *et al.*, 2012; Wu *et al.*, 2021). All expressed genes in each tissue were used as background for enrichment analysis. A GO term was considered to be an enriched term if an adjusted *p*-value < 0.05 (Benjamini-Hochberg correction). Only the top 10 GO terms were shown in the Figure.

4.2.7 Statistical analysis and visualisation

All statistical analyses were carried out using R (version 3.6.3) (Team, 2020). Plots were generated using *ggplot2* (version 3.3.2) (Wickham, 2016), *ggrepel* (version 0.8.2) (Slowikowski, 2020), *ggpubr* (version 0.4.0) (Kassambara, 2020a), and *VennDiagram* (version 1.6.20) (Chen & Boutros, 2011)

4.3 Results

4.3.1 The transcriptomic and epigenomic landscapes in mouse development and ageing

The main aim of this chapter is to investigate the relationship between gene expression and histone modification changes, including H3K4me3 representing a promoter mark and H3K27ac representing active enhancer regions during mouse development and ageing. For the prenatal mouse development, I retrieved RNA-seq and ChIP-seq (for H3K4me3 and H3K27ac) data from four tissues (forebrain, hindbrain, heart, and liver) and three developmental stages: post-conception embryonic day 11.5 (E11.5), post-conception embryonic day 15.5 (E15.5), and neonate (P0) from the ENCODE project. Likewise, for mouse ageing analysis, I used the RNA-seq and ChIP-seq (for H3K4me3 and H3K27ac) data from four corresponding tissues (olfactory bulb, cerebellum, heart, and liver) and three age groups (3, 12, and 29 months) from Benayoun *et al.* (Benayoun *et al.*, 2019). Because mouse development and ageing data come from different sources, I decided to pre-process the data from two sources using the same pipeline to make them more comparable. I downloaded raw sequencing files from both sources and pre-processed the data as described in **Section 4.2.2**.

First, I performed principal component analysis (PCA) to obtain a global view of transcriptomes across samples from ENCODE (**Figure 4.1A**) and Benayoun *et al.* data (**Figure 4.1B**). Indeed, samples are clustered by tissues in both development (**Figure 4.1A**) and ageing (**Figure 4.1B**). Samples from two brain regions, forebrain and hindbrain for development and olfactory bulb and cerebellum for ageing, are grouped together, indicating the similarity of transcriptomes in both brain regions. Interestingly, samples from the same tissue of origin were also separated by developmental stage in the analysis of ENCODE data (**Figure 4.1A**). This result was also confirmed by separately performing PCA by tissue (**Figure 4.1C**). The PCA of samples from Benayoun *et al.* by tissue also showed that samples from the same age groups tend to cluster together (**Figure 4.1D**), yet this trend was not as clear as ENCODE samples. Together, these analyses indicated that sample transcriptomes are separated by tissue, developmental stage, and age group.

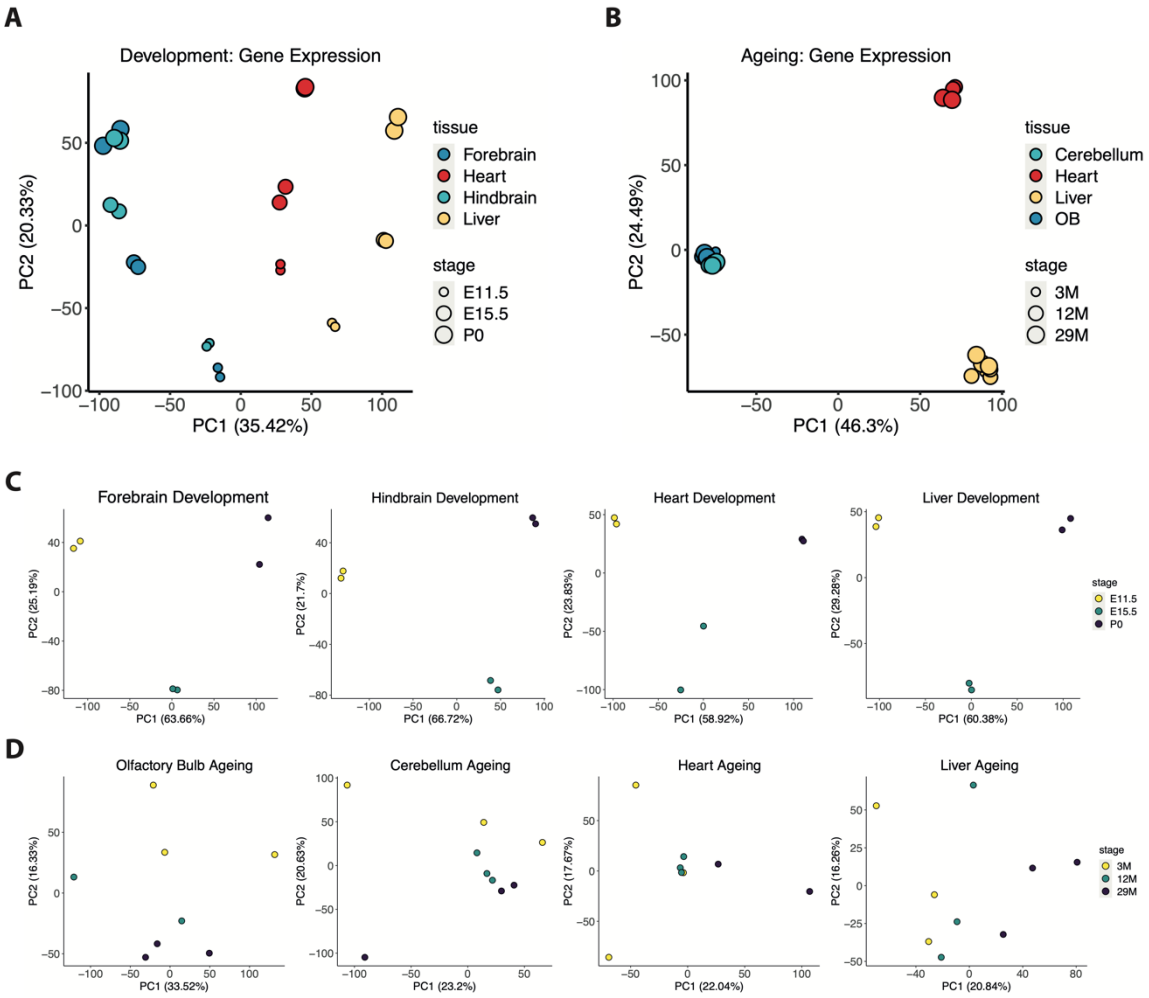


Figure 4.1. The landscape of gene expression during mouse development and ageing.
(A) PCA analysis of the gene expression during mouse development. Samples were labelled by developmental stage (E11.5, E15.5, and P0) and tissue (forebrain, hindbrain, heart, and liver). Dots' size represents the developmental stage, while dots' colours denote the tissue of origin. **(B)** PCA analysis of the gene expression during mouse ageing. Samples were labelled by developmental stage (3M, 12M, and 29M) and tissue (olfactory bulb or OB, cerebellum, heart, and liver). Dots' size represents the developmental stage, while dots' colours denote the tissue of origin. **(C)** The tissue-specific PCA analysis of the gene expression during mouse development. Colours represent developmental stages. **(D)** The tissue-specific PCA analysis of the gene expression during mouse ageing. Colours represent age groups.

Next, to ensure the reliability of the ChIP-seq pre-processing pipeline, I investigated similarities of H3K4me3 and H3K27ac ChIP-seq signals across samples. I performed pairwise correlation analyses of H3K4me3 across all tissues and developmental stages (**Figure 4.2A**). Correlation analysis revealed that samples derived from the same tissue of origin are clustered together. Samples from two brain regions, the forebrain and hindbrain, showed a high correlation in H3K4me3 signals (Pearson's correlation coefficient R = 0.8 – 0.95), indicating the similarity between promoter activity of brain tissues. The high correlation was also found between samples from heart (R = 0.82 – 0.94) and liver (R = 0.80 – 0.94). PCA confirmed three clusters corresponding to the brain (forebrain and hindbrain), heart, and liver (**Figure 4.2B**). Notably, samples from the earlier developmental stage (E11.5) of different tissues were closer together than those of later developmental stages (E14.5 and P0), suggesting that H3K4me3 patterns between tissues diverge as tissue development progress. PCA of H3K4me3 ChIP-seq data by including only samples derived from the same tissue showed clusters of samples from the same developmental stages (**Figure 4.2C**). Likewise, correlation analysis of H3K27ac signals across all samples also revealed the similarity of samples from the same tissue and developmental stage (forebrain and hindbrain: R = 0.35 – 0.83, heart: R = 0.58 – 0.83, liver: R = 0.45 – 0.78) (**Figure 4.2D**). Notably, H3K27ac signals of samples from different tissues were poorly correlated (R between the brain and other tissues = 0.0 – 0.30, R between heart and liver = 0.27 – 0.46). This observation is consistent with the idea that enhancers encode epigenetic profile of tissue specificity. The similarity of the H3K27ac pattern between samples from the same tissue and developmental stage was also confirmed by PCA (**Figure 4.2E-F**).

In a similar manner, I examined the similarity and difference of H3K4me3 and H3K27ac across samples from Benayoun *et al.* using correlation analysis and PCA (**Figure 4.3**). Indeed, H3K4me3 signals of samples from the same tissue showed a high correlation (olfactory bulb: R = 0.94 – 0.95, cerebellum: R = 0.96 – 0.97, heart: R = 0.91 – 0.94, liver: R = 0.89 – 0.92) (**Figure 4.3A**) and were grouped in PCA (**Figure 4.3B**). Similarly, H3K27ac of samples from the same tissue also showed a high correlation (olfactory bulb: R = 0.76 – 0.84, cerebellum: R = 0.91 – 0.93, heart: R = 0.85 – 0.91, liver: R = 0.81 – 0.90) (**Figure 4.3D**). Samples from the same tissues were grouped together in PCA (**Figure 4.3E**). Again, PCA of H3K4me3 and H3K27ac signals by tissue also revealed that samples could be separated by age group (**Figure 4.3C and Figure 4.3F**).

Finally, I investigated the global correlation between histone modification peak levels and gene expression. Histone modification peaks (H3K4me3 and H3K27ac) were annotated to the nearby genes. I found a significant correlation between H3K4me3 levels and gene expression in all developmental stages and age groups in all tissues (*p*-value < 2.2×10^{-16}, Pearson correlation coefficient R = 0.15 – 0.52). Likewise, a significant correlation between H3K27ac levels and gene expression was also found in all tissues (*p*-value < 2.2×10^{-16}, R = 0.11 – 0.35) (data not shown). Taken together, these analyses are consistent with the original publications (Benayoun *et al.*, 2019; Consortium *et al.*, 2020), indicating the reliability of the pre-processing pipeline.

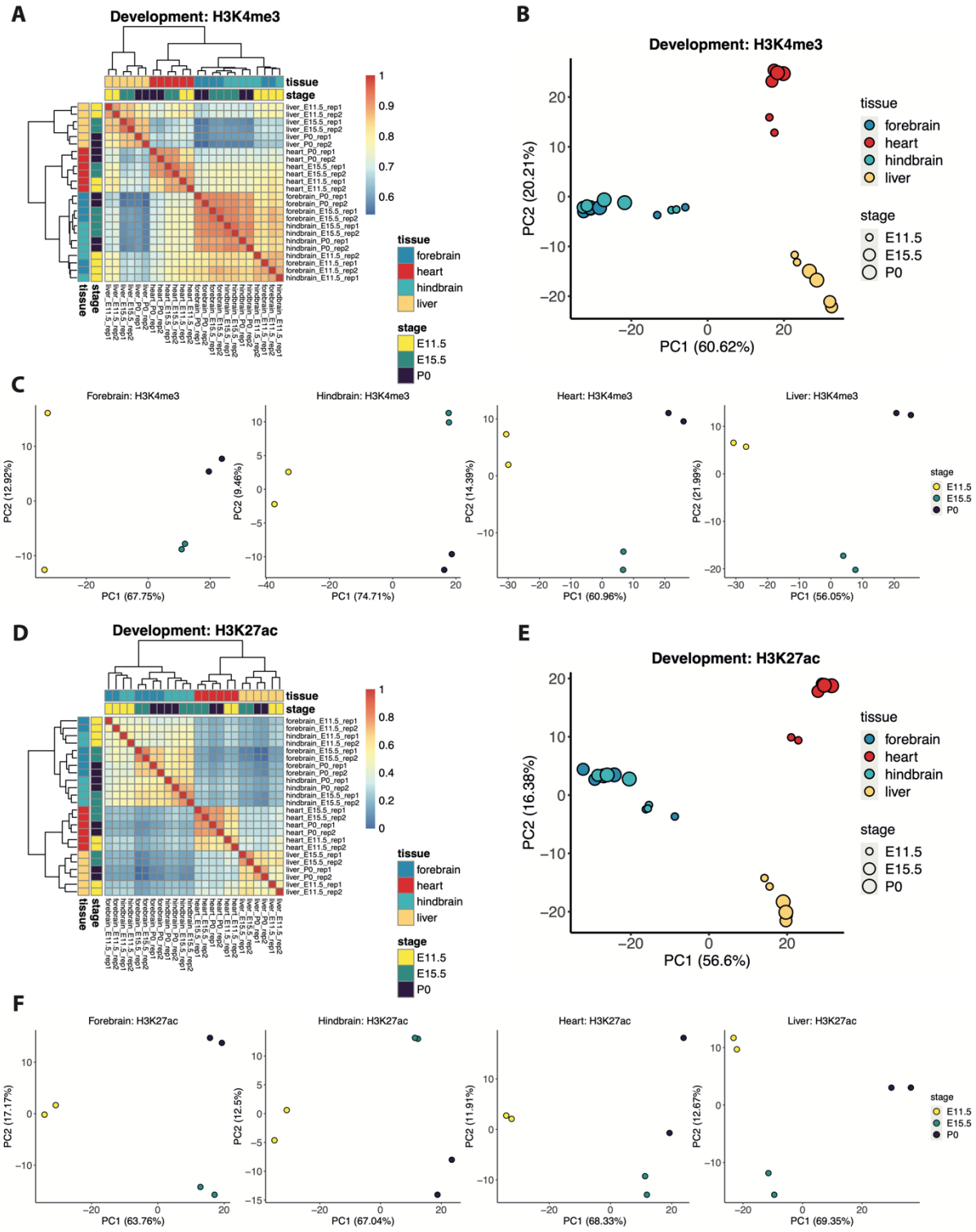

A
Development: H3K4me3
tissue
stage
forebrain
heart
hindbrain
liver
E11.5
E15.5
P0
B
Development: H3K4me3
PC2 (20.21%)
PC1 (60.62%)
C
Forebrain: H3K4me3
Hindbrain: H3K4me3
Heart: H3K4me3
Liver: H3K4me3
PC1 (67.75%)
PC2 (12.92%)
PC1 (74.71%)
PC2 (9.46%)
PC1 (60.96%)
PC2 (14.39%)
PC1 (56.05%)
PC2 (21.99%)
D
Development: H3K27ac
E
Development: H3K27ac
PC2 (16.38%)
PC1 (56.6%)
F
Forebrain: H3K27ac
Hindbrain: H3K27ac
Heart: H3K27ac
Liver: H3K27ac
PC1 (63.76%)
PC2 (17.17%)
PC1 (67.04%)
PC2 (12.5%)
PC1 (68.33%)
PC2 (11.91%)
PC1 (69.35%)
PC2 (12.67%)

Figure 4.2. The landscape of dynamic histone modifications (H3K4me3 and H3K27ac) during mouse development.

(A) Pairwise correlation of H3K4me3 signals between samples. The pairwise correlation between samples was calculated using Pearson's correlation. Samples were labelled by developmental stage (E11.5, E15.5, and P0) and tissue (forebrain, hindbrain, heart, and liver) **(B)** PCA analysis of H3K4me3 across samples. Dots' size represents the developmental stage, while dots' colours denote the tissue of origin. **(C)** PCA analysis of H3K4me3 by tissue. Colours represent developmental stages. **(D)** Pairwise correlation of H3K27ac signals between samples. The pairwise correlation between samples was calculated using Pearson's correlation. Samples were labelled by developmental stage and tissue **(E)** PCA analysis of H3K27ac across samples. Dots' size represents the developmental stage, while dots' colours denote the tissue of origin. **(F)** PCA analysis of H3K27ac by tissue. Colours represent developmental stages.

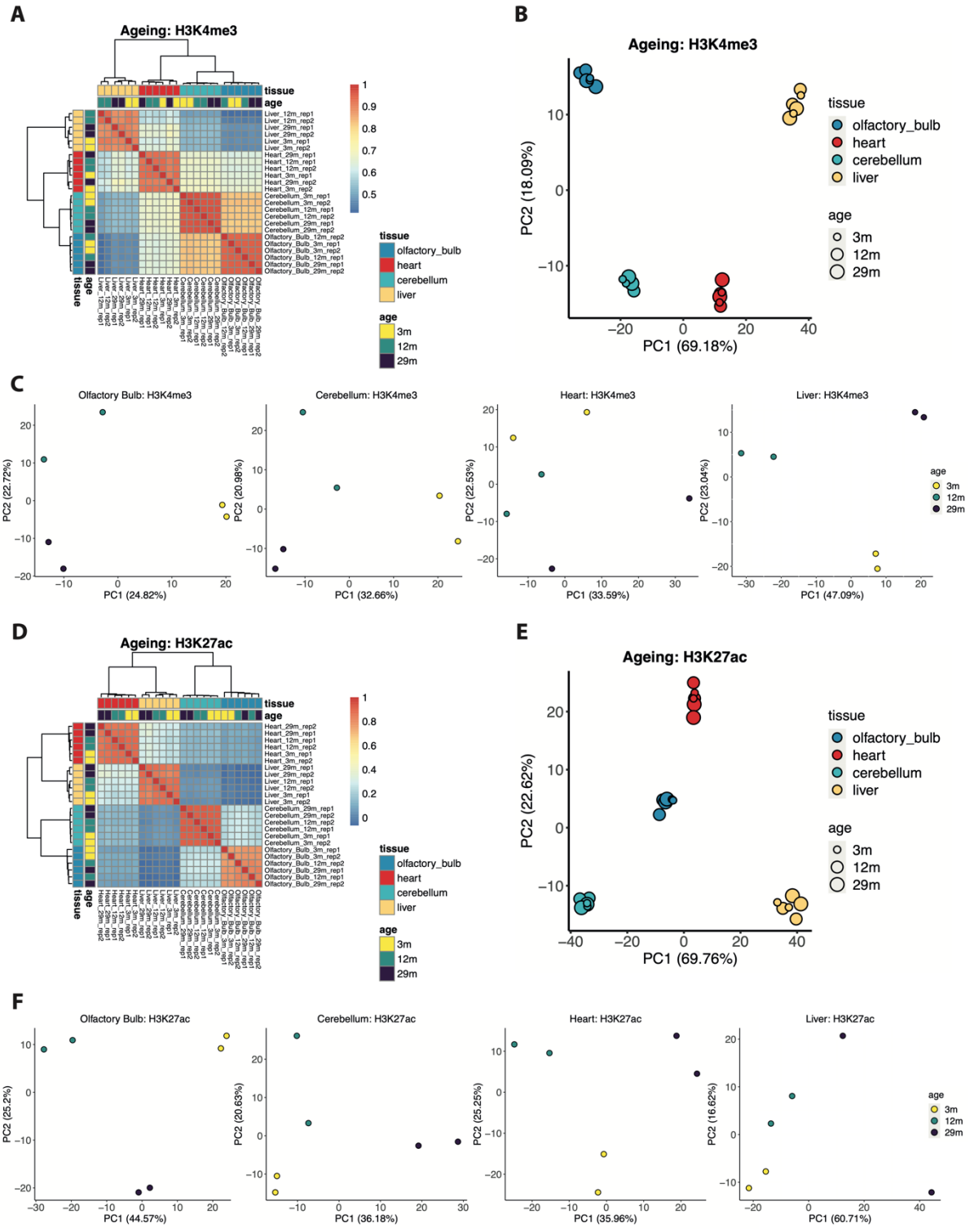

A
Ageing: H3K4me3
tissue
age
olfactory_bulb
heart
cerebellum
liver
3m
12m
29m
B
Ageing: H3K4me3
PC1 (69.18%)
PC2 (18.09%)
C
Olfactory Bulb: H3K4me3
Cerebellum: H3K4me3
Heart: H3K4me3
Liver: H3K4me3
PC1 (24.82%)
PC2 (22.72%)
PC1 (32.66%)
PC2 (20.98%)
PC1 (33.59%)
PC2 (22.53%)
PC1 (47.09%)
PC2 (23.04%)
D
Ageing: H3K27ac
E
Ageing: H3K27ac
PC1 (69.76%)
PC2 (22.62%)
F
Olfactory Bulb: H3K27ac
Cerebellum: H3K27ac
Heart: H3K27ac
Liver: H3K27ac
PC1 (44.57%)
PC2 (25.2%)
PC1 (36.18%)
PC2 (20.63%)
PC1 (35.96%)
PC2 (25.25%)
PC1 (60.71%)
PC2 (16.62%)

Figure 4.3. The landscape of dynamic histone modifications (H3K4me3 and H3K27ac) during mouse ageing.

(A) Pairwise correlation of H3K4me3 signals between samples. The pairwise correlation between samples was calculated using Pearson's correlation. Samples were labelled by the age group (3M, 12M, and 29M) and tissue (olfactory bulb, cerebellum, heart, and liver) **(B)** PCA analysis of H3K4me3 across samples. Dots' size represents age group, while dots' colours denote the tissue of origin. **(C)** PCA analysis of H3K4me3 by tissue. Colours represent age groups. **(D)** Pairwise correlation of H3K27ac signals between samples. The pairwise correlation between samples was calculated using Pearson's correlation. Samples were labelled age group and tissue **(E)** PCA analysis of H3K27ac across samples. Dots' size represents the age group, while dots' colours denote the tissue of origin. **(F)** PCA analysis of H3K27ac by tissue. Colours represent age groups.

4.3.2 The relationship between gene expression changes during development and ageing

To understand the connection between gene expression changes during development and ageing, I first separately derived tissue-specific differentially expressed (DE) genes during development and ageing. For development, I performed the pairwise differential expression analysis of samples from distinct developmental stages (E15.5 vs E11.5, P0 vs E11.5, and P0 vs E15.5) and from different age groups (12M vs 3M, 29M vs 3M, and 29M vs 12M) (**Figure 4.4A**). **Figure 4.4B-C** showed number of DE genes between conditions (adj. *p*-value < 0.05 and |FC| > 2 for development and adj. *p*-value < 0.1 and |FC| > 1.5 for ageing analyses). Indeed, the number of DE genes in development was far greater than those in ageing. This observation is not surprising, given that development involves the precise control of gene expression programmes to generate specific types of cells to support tissue functions via differentiation. On the other hand, ageing occurs in adult tissues that are already established and thus involves only subtle changes in gene expression. Genes in which significantly differentially expressed in at least one pairwise comparison in either development or ageing analyses were merged to obtain the final list of DE genes for each tissue. In total, the number of DE genes was 7063, 6082, 4856, and 6578 genes for the forebrain, hindbrain, heart, and liver, respectively (**Figure 4.5**).

Next, for each tissue, I overlapped the DE genes from development and ageing. Interestingly, a significant overlap between development and ageing DE genes was found in all four tissues (OR = 1.88 – 4.27) (**Figure 4.5**). Interestingly, about 60 – 78 per cent of ageing DE genes also changed during development (**Figure 4.5**), highlighting the connection between gene expression programme in both processes as previously suggested (Somel *et al.*, 2010) and supporting the hypothesis that some aspects of ageing might be controlled by the developmental programme (de Magalhaes, 2012).

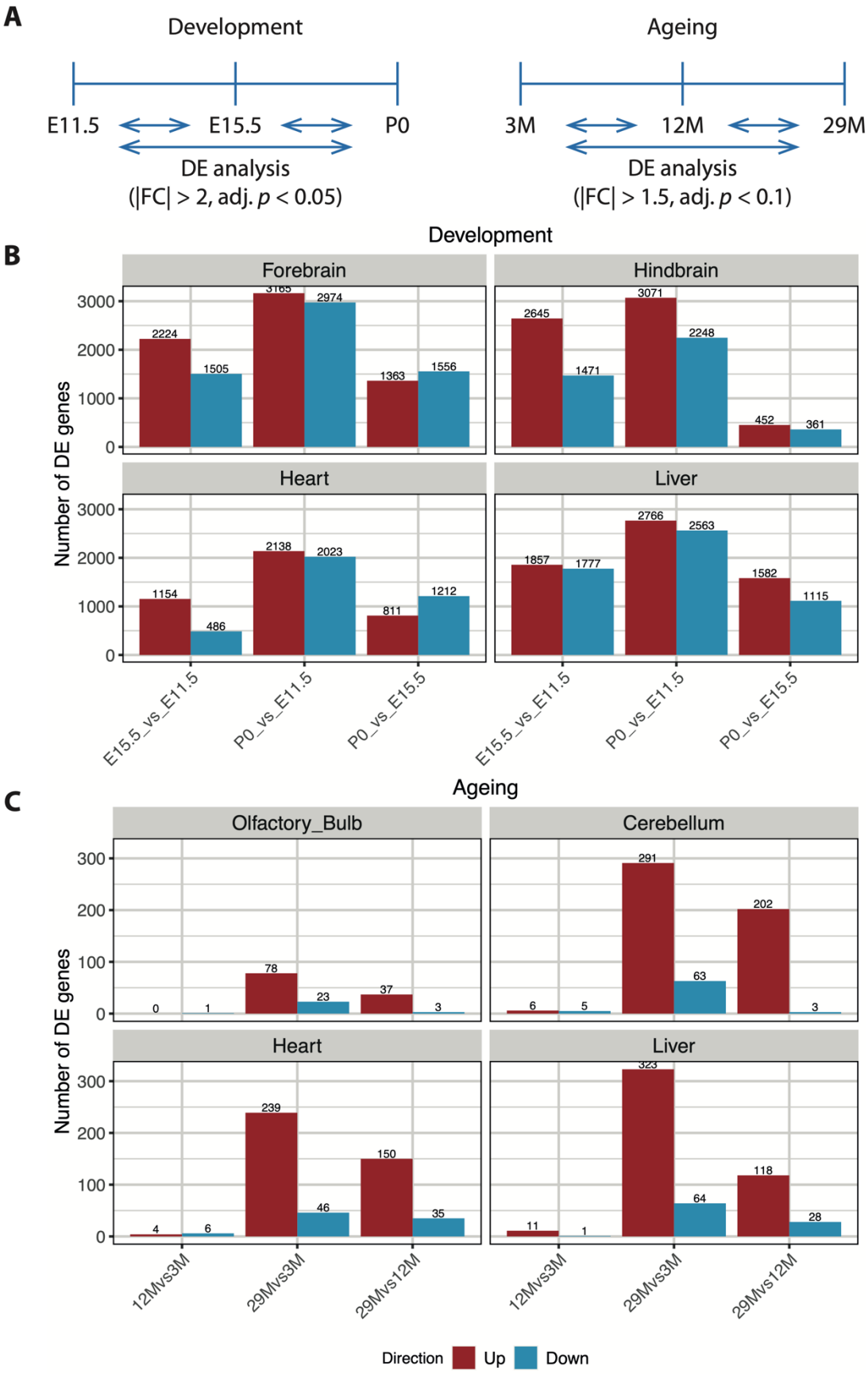

A
Development
E11.5
E15.5
P0
DE analysis
(|FC| > 2, adj. p < 0.05)
Ageing
3M
12M
29M
DE analysis
(|FC| > 1.5, adj. p < 0.1)
B
Development
Forebrain
Hindbrain
Heart
Liver
Number of DE genes
E15.5_vs_E11.5
P0_vs_E11.5
P0_vs_E15.5
C
Ageing
Olfactory_Bulb
Cerebellum
Heart
Liver
12Mvs3M
29Mvs3M
29Mvs12M
Direction
Up
Down

Figure 4.4. Differentially expressed (DE) genes in development and ageing.

(A) The diagram shows the pairwise comparison of gene expression between development stages and age groups. **(B)** Bar chart shows the number of significant DE genes between developmental stages in each tissue (adj. *p*-value < 0.05 and |FC| > 2). **(C)** Bar chart shows the number of significant DE genes between age groups in each tissue (adj. *p*-value < 0.1 and |FC| > 1.5). The number of downregulated genes was shown in blue bars, whilst the number of upregulated genes was shown in red bars.

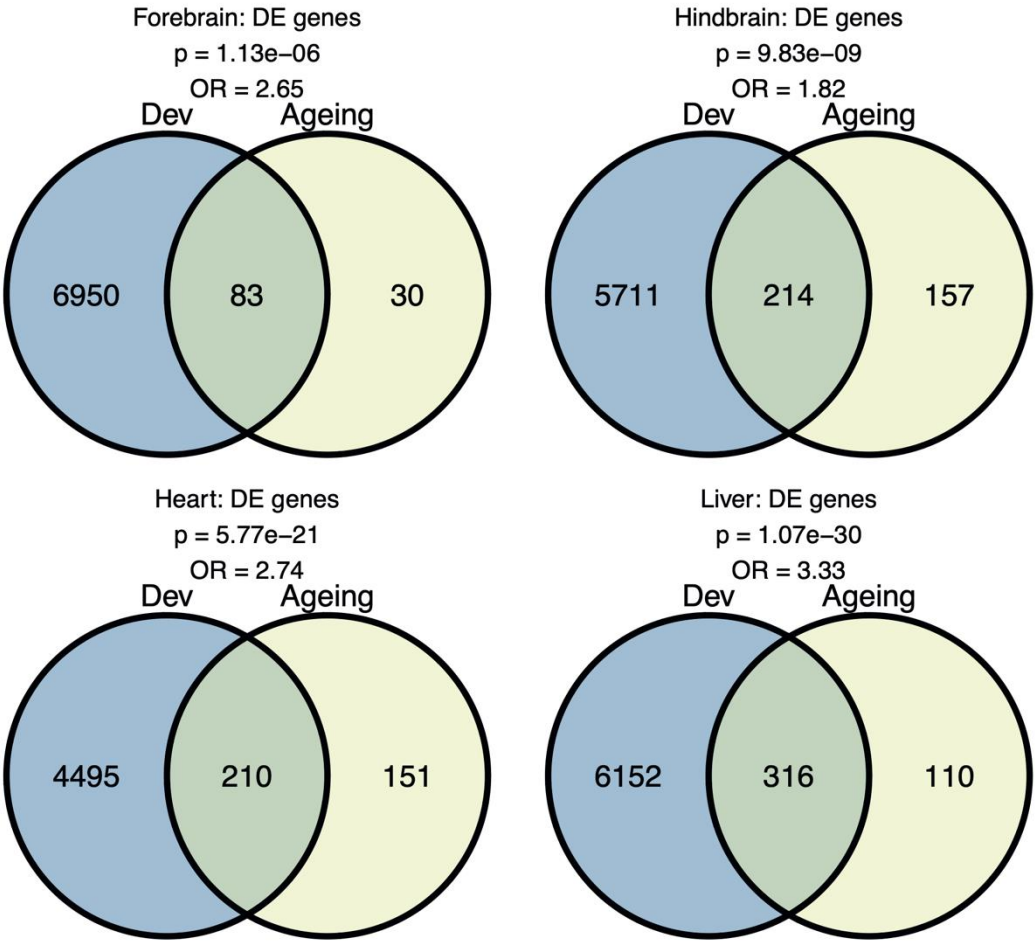
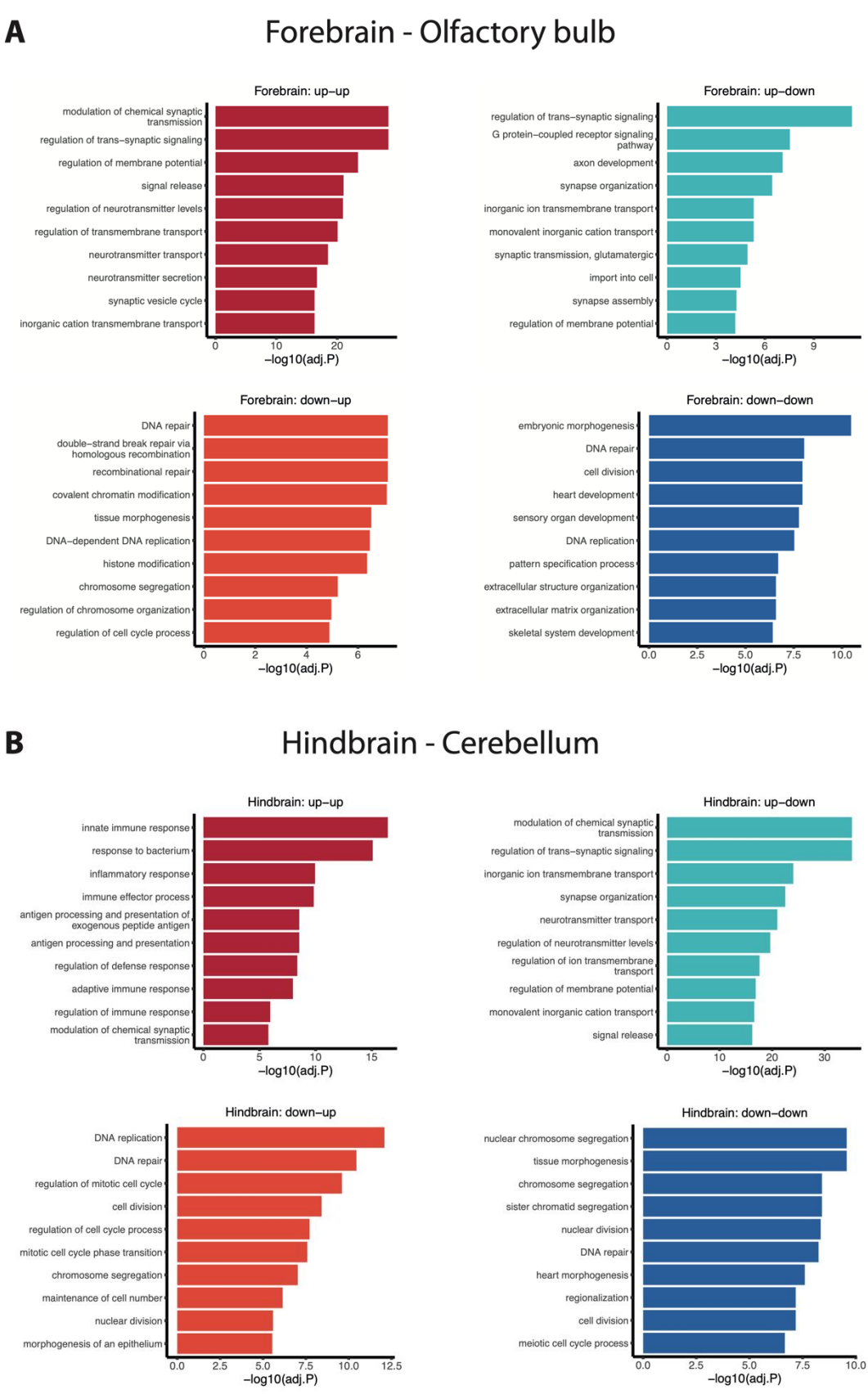


Figure 4.5. Overlap of development and ageing differentially expressed (DE) genes.

The overlap analysis between DE genes in development and ageing for each tissue was performed using Fisher's exact test. Venn diagrams represent the number of DE genes in development, ageing, and shared genes between both processes. *P*-values and odds ratios (OR) from the Fisher's exact tests were shown above the diagrams.

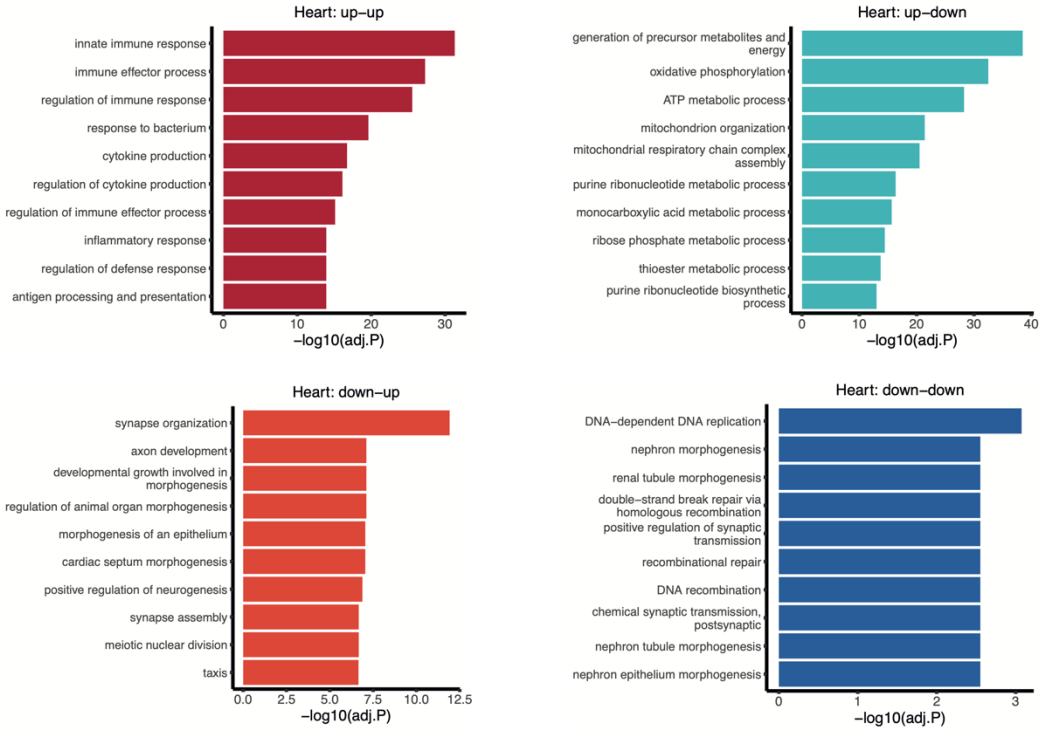
I next classified total DE genes (the combined list of DE genes in development and ageing) for each tissue into different trajectories. Genes whose expression level is higher in P0 than E11.5 are classified as those upregulated during development, while those with a lower expression level in P0 than E11.5 are downregulated genes. Likewise, upregulated and downregulated genes in ageing refer to genes whose expression in 29M is higher or lower than in 3M, respectively. Therefore, total DE genes for each tissue were classified into four groups (development-ageing) as follows: up-up, up-down, down-up, and down-down.

To better understand the biological functions of DE genes in development and ageing, I performed GO analysis on genes in four classified groups for each tissue (**Figure 4.6**). This analysis revealed several features of genes in each group. For example, the up-up genes were associated with immune system functions in most tissues except the forebrain, suggesting the immune system maturation during development and consistent with the immune cell infiltration in ageing tissues (Mogilenko *et al.*, 2021). The cell cycle and DNA replication genes were enriched in down-up and down-down groups in several tissues, potentially indicating reduced proliferation in late developmental states. Notably, DNA repair genes are reduced both during development and ageing. These genes may be highly expressed early in development to prevent DNA damage accumulation during rapid cell division. Their expression might decrease as the cell cycle slows down in late development. Interestingly, I observed enrichment of tissue-specific functions in up-down genes in all tissues, such as synapses in brain, mitochondria and metabolic processes in heart, and metabolism in liver. This result might point toward the establishment of specialised cellular functions during development and the loss of cellular and tissue identity during ageing.


A
Forebrain - Olfactory bulb
Forebrain: up–up
modulation of chemical synaptic transmission
regulation of trans–synaptic signaling
regulation of membrane potential
signal release
regulation of neurotransmitter levels
regulation of transmembrane transport
neurotransmitter transport
neurotransmitter secretion
synaptic vesicle cycle
inorganic cation transmembrane transport
0 10 20
–log10(adj.P)
Forebrain: up–down
regulation of trans–synaptic signaling
G protein–coupled receptor signaling pathway
axon development
synapse organization
inorganic ion transmembrane transport
monovalent inorganic cation transport
synaptic transmission, glutamatergic
import into cell
synapse assembly
regulation of membrane potential
0 3 6 9
–log10(adj.P)
Forebrain: down–up
DNA repair
double–strand break repair via homologous recombination
recombinational repair
covalent chromatin modification
tissue morphogenesis
DNA–dependent DNA replication
histone modification
chromosome segregation
regulation of chromosome organization
regulation of cell cycle process
0 2 4 6
–log10(adj.P)
Forebrain: down–down
embryonic morphogenesis
DNA repair
cell division
heart development
sensory organ development
DNA replication
pattern specification process
extracellular structure organization
extracellular matrix organization
skeletal system development
0.0 2.5 5.0 7.5 10.0
–log10(adj.P)
B
Hindbrain - Cerebellum
Hindbrain: up–up
innate immune response
response to bacterium
inflammatory response
immune effector process
antigen processing and presentation of exogenous peptide antigen
antigen processing and presentation
regulation of defense response
adaptive immune response
regulation of immune response
modulation of chemical synaptic transmission
0 5 10 15
–log10(adj.P)
Hindbrain: up–down
modulation of chemical synaptic transmission
regulation of trans–synaptic signaling
inorganic ion transmembrane transport
synapse organization
neurotransmitter transport
regulation of neurotransmitter levels
regulation of ion transmembrane transport
regulation of membrane potential
monovalent inorganic cation transport
signal release
0 10 20 30
–log10(adj.P)
Hindbrain: down–up
DNA replication
DNA repair
regulation of mitotic cell cycle
cell division
regulation of cell cycle process
mitotic cell cycle phase transition
chromosome segregation
maintenance of cell number
nuclear division
morphogenesis of an epithelium
0.0 2.5 5.0 7.5 10.0 12.5
–log10(adj.P)
Hindbrain: down–down
nuclear chromosome segregation
tissue morphogenesis
chromosome segregation
sister chromatid segregation
nuclear division
DNA repair
heart morphogenesis
regionalization
cell division
meiotic cell cycle process
0.0 2.5 5.0 7.5 10.0
–log10(adj.P)

C

Heart



D

Liver

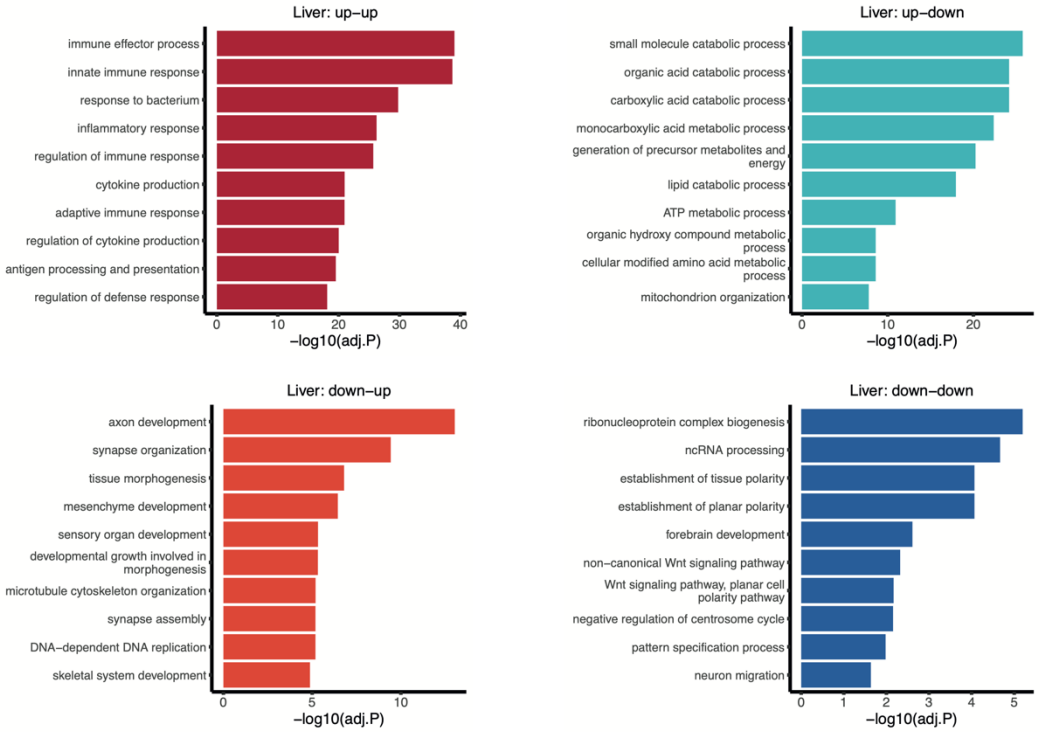


Figure 4.6. Functional enrichment analysis of development and ageing DE genes with diverse trajectories in the forebrain (A), hindbrain (B), heart (C), and liver (D).

The total DE genes were classified into four groups in each tissue based on the direction of expression changes in development and ageing. GO enrichment analysis was performed in each group of genes. A GO term is considered a significantly enriched term if adj. *p*-value < 0.05. The top 10 enriched terms were extracted and plotted in the graph.

4.3.3 Alterations of tissue-specific gene expression during development and ageing

A recent study suggested that development is related to the establishment of tissue and cellular identities and shows an inter-tissue divergence pattern. On the other hand, ageing is associated with the loss of tissue specificity. Thus, gene expression changes between tissues tend to converge towards similar levels. Therefore, the authors proposed that inter-tissue gene expression changes follow the divergence-convergence (DiCo) pattern across the lifespan (Izgi *et al.*, 2022). Here, I further explored this hypothesis by examining whether the DiCo pattern also happens in prenatal development and ageing using ENCODE and Benayoun *et al.* datasets. For each gene, I calculated the coefficient of variation (CoV), the ratio of standard variation to the mean, to measure the gene expression variability between tissues (brain, heart, and liver) in relation to the mean in each developmental stage or age group. Indeed, CoV between tissues increases as development progresses (**Figure 4.7A**), consistent with the idea that development involves the divergence of gene expression between tissues. Interestingly, while I did not observe the difference between CoV from 3M and 12M (adj. *p*-value = 0.273), the CoV between tissue showed a slight yet significant decreased in the 29M group comparing with 3M and 12M (adj. *p*-value = 3.88×10^{-13} and 1.23×10^{-16}, respectively) (**Figure 4.7B**). Therefore, this result further strengthens the hypothesis that mammalian tissue transcriptomes diverge from each other during development, yet they converge during ageing, as proposed by Izgi *et al.* (Izgi *et al.*, 2022).

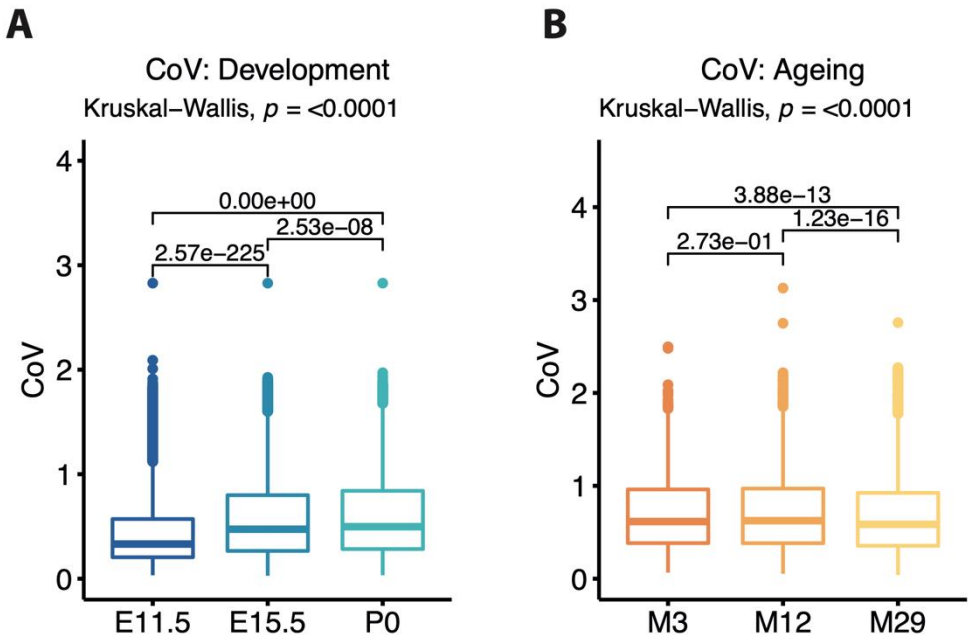


Figure 4.7. Coefficient of variation (CoV) in development (A) and ageing (B).
For each gene that is expressed in all three tissues (brain, heart, and liver), CoV was calculated as the ratio between standard deviation over mean expression. The group comparison was performed by the Kruskal-Wallis test. The pairwise comparisons were performed using the two-sided Dunn's test. *P*-values from Dunn's test adjusted by Bonferroni correction are shown.

Next, I reasoned that if development is related to the divergence of tissue gene expression while ageing involves the loss of tissue specificity, tissue-specific genes should increase during development but decrease during ageing. To further explore this hypothesis, I identified tissue-specific genes in each developmental stage or age group by performing differential expression analysis comparing the tissue of interest and other tissues. For example, brain-specific genes were derived from the comparison between brain samples (forebrain and hindbrain) and samples from other tissues (heart and liver) (adj. p-value < 0.01 and |FC| > 5). Likewise, liver-specific genes were differentially expressed genes in the liver compared with other tissues (forebrain, hindbrain, and heart). The numbers of tissue-specific genes increase as the development stage progresses, supporting the establishment of tissue identity (**Figure 4.8A**). For ageing, I observed that the number of tissue-specific genes only decreased slightly with age (**Figure 4.8B**). Thus, this analysis supported the establishment of the tissue identity during prenatal development.

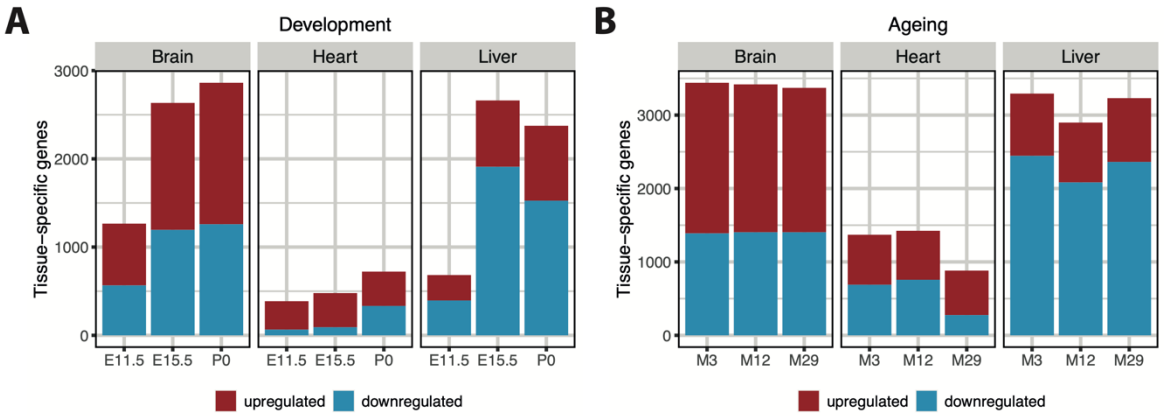


Figure 4.8. The number of tissue-specific genes in development (A) and ageing (B). In each developmental stage or age group, differential expression analysis was performed by comparing the expression of tissue of interest (e.g., brain) against other tissues (e.g., heart and liver). Number of significant upregulated and downregulated genes (adj. *p*-value < 0.01 and |FC| > 5) were plotted in red and blue, respectively.

Next, to obtain a single set of tissue-specific genes for each tissue (here referred to as "confident tissue-specific genes"), I retained only upregulated tissue-specific genes that were shared between at least two developmental stages or age groups. In total, confident tissue-specific genes for brain, heart and liver were 2241, 781, and 1165 genes, respectively. I further investigated changes in gene expression of these confident tissue-specific genes during development and ageing. While tissue-specific genes increased by development stages (**Figure 4.9A**), these genes did not show a clear increase or decrease patterns during ageing (**Figure 4.9B**).

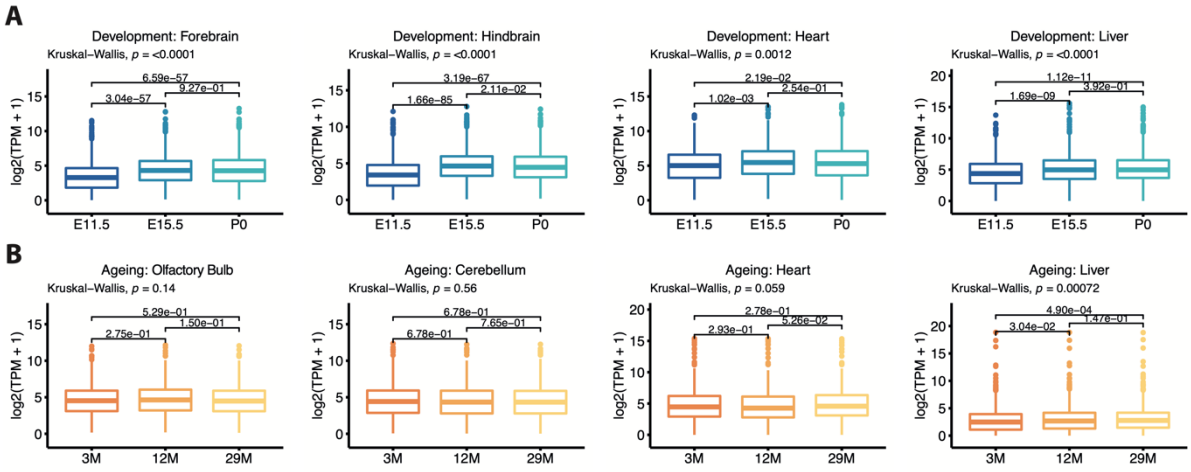


Figure 4.9. Changes in tissue-specific gene expression level during development and ageing.

The differences in the expression of confident tissue specific genes between developmental stages **(A)** and between age groups **(B)** were investigated. The group comparison was performed by the Kruskal-Wallis test. The pairwise comparisons were performed using the two-sided Dunn's test. *P*-values from Dunn's test between developmental stages/age groups adjusted by Bonferroni correction are shown.

I further investigated potential changes in the activity of promoters and enhancers that are located close to confident tissue-specific genes during prenatal development and ageing. Indeed, H3K4me3 peaks of tissue-specific genes showed an increasing trend in mouse development in most tissues except the heart (**Figure 4.10A**). Likewise, H3K27ac of tissue-specific genes also increased during developmental stages in all four tissues (**Figure 4.10B**). Together, these results further highlighted the establishment of the tissue identity during prenatal development at an epigenome level. For ageing, however, I cannot observe a clear trend of changes in H3K4me3 and H3K27ac peaks associated with tissue-specific genes (**Figure 4.10A-B**).

In summary, this analysis revealed the inter-tissue divergence in gene expression during development (**Figure 4.7A**), which was partly driven by the expression and chromatin landscape associated with tissue-specific genes (**Figure 4.8-4.10**). Conversely, while there was a trend toward the loss of tissue specificity during ageing (**Figure 4.7B**), the contribution of tissue-specific genes and histone marks associated with these genes were less clear (**Figure 4.8-4.10**). One potential explanation is that changes in the ageing gene expression programme

are usually subtle. Therefore, it is difficult to observe such contribution from tissue-specific genes in ageing with only a limited sample size in this study.

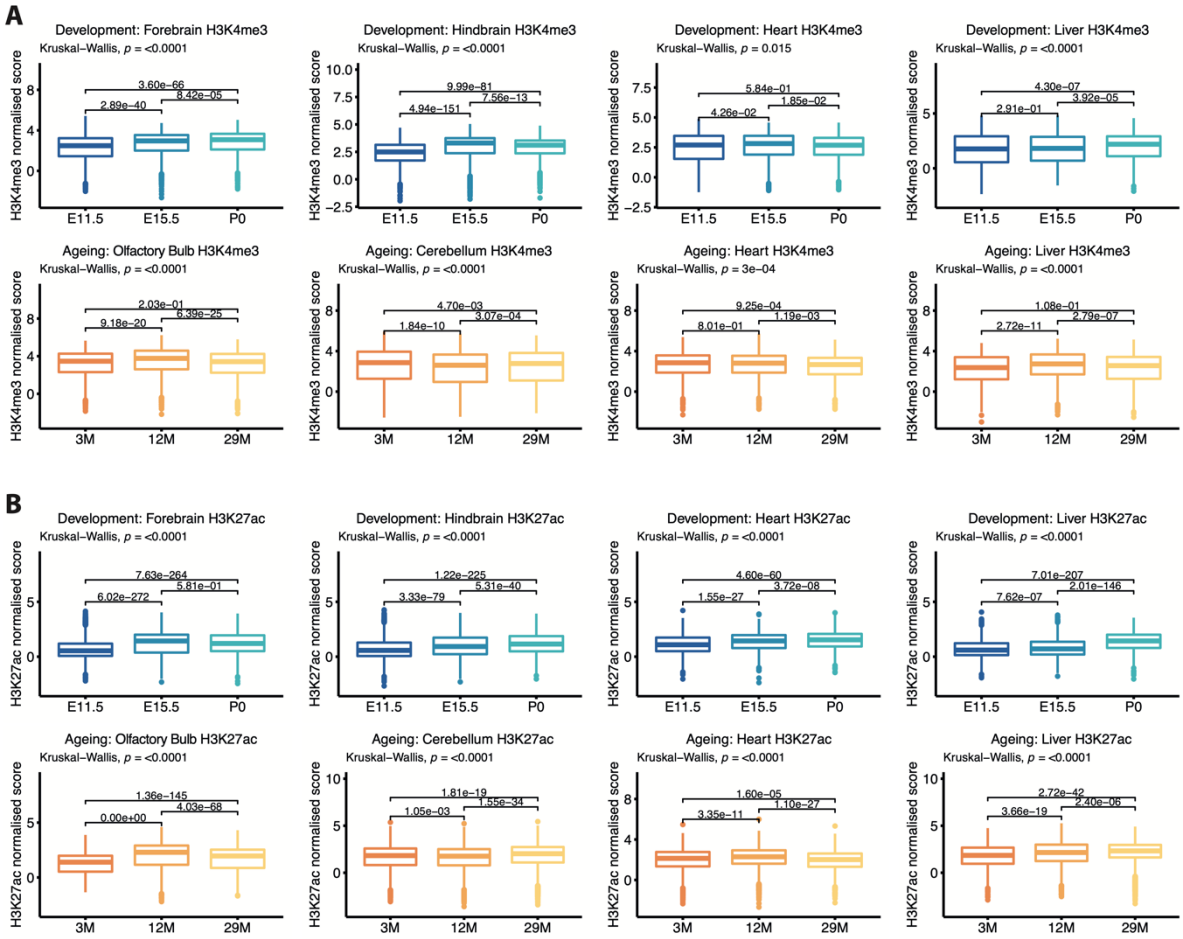


Figure 4.10. Changes in histone modification marks that are associated with tissue-specific genes during development and ageing.

The differences in H3K4me3 **(A)** and H3K27ac **(B)** levels of peaks located nearby the confident tissue specific genes during development and ageing were investigated. The group comparison was performed by the Kruskal-Wallis test. The pairwise comparisons were performed using the two-sided Dunn's test. *P*-values from Dunn's test between developmental stages/age groups adjusted by Bonferroni correction are shown.

4.4 Discussion

Several lines of evidence suggest the potential relationship between postnatal development and ageing in gene expression levels (Somel *et al.*, 2010; Donertas *et al.*, 2017; Izgi *et al.*, 2022). However, it remains unclear if the transcriptomic changes during prenatal development also relate to those changes later in life. In addition, the contribution of epigenetic processes in regulating development and ageing transcriptomes has not been studied. To bridge these knowledge gaps, I analysed the available gene expression and histone modification data from prenatal mouse development and ageing across four tissues (forebrain, hindbrain, heart, and liver). I confirmed that changes in gene expression during ageing partly originated in prenatal development.

4.4.1 Dynamic gene expression across life course

In this chapter, I showed that differentially expressed genes in prenatal development and ageing significantly overlapped in all tissues tested. This result extends the findings from previous studies reporting that changes in gene expression during ageing partly originated from postnatal development in several tissues, such as the liver, lung, kidney, and brain (Lui *et al.*, 2010; Somel *et al.*, 2010). Importantly, I classified differentially expressed genes from trajectories of changes in development and ageing and identified common biological processes that change in each trajectory across tissues. For instance, genes involved in the cell cycle tend to decrease during development and ageing across several tissues (down-down). This result is not surprising, given that cell proliferation decreases as the developmental stage progresses. In addition, ageing is associated with the loss of stem cell functions, therefore, decreased expression of cell cycle genes was observed in several studies (de Magalhaes *et al.*, 2009; Chatsirisupachai *et al.*, 2019). On the other hand, immune-related processes increase during development and ageing (up-up). Indeed, this result could reflect the development of tissue-resident immune cells, such as macrophages and microglia cells in the brain (La Manno *et al.*, 2021). Immune cell infiltration and inflammation that occur during ageing could explain an increase in immune-related terms (Mogilenko *et al.*, 2021; Palmer *et al.*, 2021). Furthermore, genes involved in tissue-specific functions, such as synapses in brain and metabolism in liver, are enriched in genes that show reversal patterns between development and ageing (up-down and down-up). Consistent with this observation, a study by Izgi *et al.* also revealed that genes

exhibiting a reversal pattern are not shared among tissues and tend to be tissue-specific (Izgi *et al.*, 2022).

4.4.2 The contribution of tissue-specific genes in the divergence-convergence gene expression pattern

As previously reported (Izgi *et al.*, 2022), I confirmed the overall trend that gene expression among tissues diverges during development and tends to converge during ageing by using CoV analysis. In addition, an increase in the number of tissue-specific genes during prenatal development further confirmed a divergence pattern of gene expression. In ageing, however, the decrease in the number of tissue-specific genes with age was not very clear. Likewise, I found an increase in the expression of tissue-specific genes as the developmental stage progresses. Yet, the expression level of tissue-specific genes did not show an obvious decreasing or increasing trend during ageing. I extended this analysis to the epigenetic level by investigating the dynamic levels of H3K4me3 and H3K27ac peaks that were annotated to tissue-specific genes. Consistent with the gene expression, H3K4me3 and H3K27ac peaks of tissue-specific genes increased with the developmental stage but did not show a clear trend during ageing.

Notably, the data I used in this chapter come from three discrete age groups (3 months, 12 months, and 29 months). The study by Izgi *et al.* contained time-series gene expression data from seven mice for the postnatal development series (2 to 61 days) and nine mice for the ageing series (93 to 904 days) (Izgi *et al.*, 2022). They used age as a continuous variable in the analysis, such as by performing a correlation analysis between age and gene expression level. Thus, this fundamental difference in the data analysis may prevent the analysis in this chapter from capturing the convergence pattern that was observed in the Izgi *et al.* study. Therefore, future studies with a higher sample size spanning larger age groups would be necessary to further validate the DiCo pattern in gene expression and histone modification during development and ageing.

Another important point is that the analyses in this chapter were done using “bulk” data. Thus, it is unclear if the observed patterns arise from cellular composition or cell intrinsic changes. The study by Izgi *et al.* has used single-cell RNA-sequencing data to show that cell type composition changes during development and ageing may partly explain the DiCo pattern (Izgi

et al., 2022). For example, hepatocytes in liver increase in frequency during development but decrease during ageing, supporting the establishment and loss of tissue specificity during development and ageing, respectively. However, they also found increased expression similarity between distinct cell types with age, suggesting that cell-autonomous gene expression changes could also contribute to the DiCo pattern. It is possible that epigenetic changes with ageing could contribute to cell-autonomous changes in gene expression (Hernando-Herraez *et al.*, 2019). However, single-cell multi-omics datasets are required to explain the role of age-related epigenetic shifts in modulating gene expression at a single-cell level.

4.4.3 Future questions and analyses

Throughout this chapter, I performed the analysis in a gene expression centric manner. The work in this chapter is, however, not completed. It is also interesting to identify differential enriched H3K4me3 and H3K27ac peaks during development and ageing. We could ask if there is an overlap between dynamic epigenome regulation during development and ageing by performing such an analysis. Furthermore, it would be interesting to analyse the dynamic changes in H3K4me3 and H3K27ac during development and ageing and classify them into different trajectories, like what was done at the gene expression level (**Figure 4.6**). Such analysis could further reveal if the epigenetic changes correspond to the observed patterns found from gene expression analysis, as already discussed in **section 4.4.1**.

In addition, active histone marks such as H3K4me3 and H3K27ac data will also allow the identification of common transcription factors that potentially regulate gene expression in both development and ageing. For example, a previous study on *C. elegans* has identified transcription factors that could control both development and ageing by using chromatin accessibility data across the lifespan (Janes *et al.*, 2018). Indeed, the relationship between development and ageing in the epigenome is far less studied compared with the transcriptome level. Apart from H3K4me3 and H3K27ac, other epigenetic components, such as DNA methylation, chromatin accessibility, and other histone modification marks, are also important in gene regulation. Thus, further studies and new bulk and single-cell datasets are needed to better elucidate how dynamic epigenomic changes during development relate to those changes during ageing.

Chapter 5

Concluding remarks

5.1 Conclusions and discussion

This thesis leveraged publicly available 'omic' datasets and a computational genomics approach to investigate the relationship between ageing, cancer, and cellular senescence (**Chapter 2**). This study also explored the differences in cancer molecular landscape as a function of the patient's age (**Chapter 3**). In addition, the present study sought to better understand the relationship between mammalian development and ageing by using mice transcriptomic and epigenetic datasets (**Chapter 4**). The main results of the thesis can be summarised as follows:

1. The relationship between ageing and cancer gene expression is tissue specific. In most tissue, ageing and cancer gene expression patterns change in opposite directions. On the other hand, ageing and cancer genes are altered in the same direction in the thyroid and uterus.
2. Cellular senescence signatures, genes that are consistently overexpressed or underexpressed in cellular senescence across datasets, were identified by meta-analysis using 20 microarray datasets. This gene list serves as a resource for researchers interested in cellular senescence genes.
3. Cellular senescence signatures changed in the same direction as ageing genes but were strongly opposed to cancer genes.
4. Tumours from older patients present an overall increase in genomic instability, somatic copy-number alterations (SCNAs), and somatic mutations.
5. Several SCNA regions, both arm-level and focal-level, were found to be exhibited age-associated patterns, particularly in low-grade glioma, endometrial cancer, and ovarian cancer. Some SCNA regions were altered more frequently in younger patients, while others were gained or lost more often in older patients. Several cancer driver genes were located in age-associated SCNA regions.
6. Several cancer driver genes exhibited age-associated mutational patterns. These age-associated driver genes tend to be cancer-specific. The SCNA and mutational analyses

provide a comprehensive view of age-associated genomic alterations across cancers. These results are expected to be a valuable resource for research communities.

7. The age-associated genomic alterations were potentially related to age-related cancer subtypes, such as the IDH-mut subtype in gliomas and the hypermutation subtype in endometrial cancer.
8. Age-related transcriptomic changes are partly regulated by age-related DNA methylation changes and associated with several biological processes.
9. Genes exhibiting expression changes during mammalian development and ageing significant overlapped across tissues.
10. Transcriptomes across tissues diverge during development and tend to converge during ageing, supporting the previously reported divergence-convergence behaviour. This pattern partly emerged due to increased expression and active chromatin landscape of tissue-specific genes during development. However, the contribution of expression and chromatin landscape of tissue-specific genes in ageing is still unclear.

Together, these findings advance our understanding of the complex relationship between transcriptomic changes in human ageing, cancer, and cellular senescence. The findings from this thesis also provide a comprehensive view of how age impacts cancer genotypes, which will be valuable in personalised medicine. Furthermore, they suggest a relationship between mammalian development and ageing, a topic worth further study. Importantly, several sets of results from the analyses in this thesis could serve as a resource for further studies, such as tissue-specific differentially expressed genes in human ageing, cellular senescence signatures, and a list of age-associated cancer driver genes. Below, I discussed some of the topics related to the above-mentioned findings. Some parts of this discussion were included in the accepted review article in *Trends in Cancer* entitled "Age-associated differences in the cancer molecular landscape".

5.1.1 A computational genomic approach to study the relationship between biological processes across life course and the current limitations

Throughout this thesis, I performed various computational genomic analyses using several publicly available datasets, including those generated by international consortium efforts such as TCGA, GTEx, and ENCODE. Thus, this reiterates the importance of open data sharing, which has been and will continue to benefit research communities. The availability of such

large-scale datasets opens an opportunity to perform several types of analyses. For example, gene expression data from GTEx comes from numerous tissues/organs, allowing the identification of age-related differentially expressed genes in a tissue-specific manner and the comparison across different tissues. I also used gene expression data from several microarray datasets to generate cellular senescence signatures. This approach helps increase the reliability of differentially expressed genes identified from individual datasets.

In **Chapter 2**, I investigated the relationship between ageing and cancer gene expression using RNA-seq data from GTEx (Consortium, 2013) and TCGA (Cancer Genome Atlas Research *et al.*, 2013b). These gene expression data, however, are derived from bulk samples. Thus, the observed gene expression level is an average gene expression from millions of cells from different cell types within a tissue. One of the significant disadvantages is that the bulk assay cannot distinguish whether observed changes in gene expression with age come from changes in cell-type composition within a tissue or from cell-type-specific transcriptomic changes. Several recent studies employed scRNA-seq to examine age-related transcriptomic changes in mice and humans to resolve this issue. For example, the Tabula Muris Consortium has generated the 'Mouse Ageing Cell Atlas' by using scRNA-seq across the lifespan of mice, including data from 23 tissues and organs (Tabula Muris, 2020). This analysis revealed changes in cellular composition during ageing, such as increased plasma cell percentage and decreased T cells with age in the spleen. Furthermore, this single-cell analysis also revealed that the fractions of cells expressing *Cdkn2a* increase with age, suggesting age-related accumulation of senescent cells (Tabula Muris, 2020). In addition, another recent study on the mouse mammary gland revealed age-related changes in cell proportions and gene expression (Li *et al.*, 2020a). These age-associated alterations could facilitate tumour progression, such as loss of extracellular matrix (ECM) integrity, compromised endothelial barrier, and increased production of proinflammatory cytokines. However, the sample size of these scRNA-seq studies is still limited to only a few individuals. Further development of single-cell techniques will facilitate our understanding of the relationship between ageing and cancer.

For the analysis of age-associated differences in cancer molecular landscape, I used pan-cancer multi-omic datasets from TCGA, which comprise ~10,000 tumours from 33 cancer types (Cancer Genome Atlas Research *et al.*, 2013b). By analysing the tumour molecular landscape on the age of the patients in these pan-cancer datasets (**Chapter 3**), I was able to show that some cancer had a robust age-associated pattern, such as endometrial cancer and gliomas. On

the other hand, many cancer types did not display clear differences between tumours of different ages. For example, I did not observe any significant mutation in the cancer driver gene and differentially expressed gene by age in pancreatic cancer. The distinct age-related pattern across cancer types can only be investigated using this large-scale resource. Indeed, other pan-cancer datasets, such as the Pan-Cancer Analysis of Whole Genome (PCAWG) (Consortium, 2020) and Genomics Evidence Neoplasia Information Exchange (AACR GENIE) (Consortium, 2017), have also been used recently to study age-associated differences in molecular landscape (Li *et al.*, 2022). However, several studies on the topic of age-associated differences in cancer molecular landscape relied heavily on only a few datasets, such as TCGA, AACR GENIE, and METABRIC (Curtis *et al.*, 2012). Moreover, current datasets usually have limited numbers of samples from extreme age groups, particularly for 20-30 and 80-90 years old, obscuring the findings of the molecular alterations specific to these age ranges. Therefore, there is an urgent need for new pan-cancer datasets to validate the findings from our and others and discover novel information.

Throughout the study in **Chapter 3**, I included age as a continuous variable. I used statistical methods such as linear and logistic regression to study the association between age and variables of interest (e.g. somatic mutations, copy-number alterations, and gene expression). Numerous studies, however, defined age cut-offs to distinguish between young and old groups. For instance, Liao *et al.* studied differences between premenopausal and postmenopausal breast cancer using ≤45 and ≥55 years old as age cut-offs, respectively (Liao *et al.*, 2015). Another study investigated genomic landscape differences between very young (≤35) and old breast cancer (≥45) (Waks *et al.*, 2022). Thus, the age cut-off varies between studies. This lack of consistency regarding age cut-offs between studies could impair the reproducibility of the findings. Furthermore, ageing may happen at different rates in different organs. For these reasons, it is not obvious to define a single age cut-off to distinguish between young and old groups and a continuous age variable is preferred in the thesis.

Chapter 4 investigated the relationship between transcriptomic and epigenetic changes during mammalian development and ageing. In this chapter, I used paired transcriptome and epigenome data from ENCODE project (Consortium *et al.*, 2020) and another published dataset (Benayoun *et al.*, 2019). The availability of several 'omic' layers within the same dataset, such as gene expression and epigenetics, allows the integration of different data types. However, as the analyses in the chapter are not completed, further analyses can be performed, such as

dynamic histone modification across the lifespan and transcription factor motif analysis of potential transcription factors that regulate both development and ageing. How histone modifications affect gene expression is unclear and is currently an active research area. For instance, histone modification levels can accurately predict gene expression (Karlic *et al.*, 2010). However, a more recent study reported that blocking transcription removed H3K4me3 and H3K27ac from chromatin, indicating that transcription is required for histone modifications (Wang *et al.*, 2022). Therefore, the mechanistic relationship between different histone modifications and gene expression, particularly during development and ageing, remains to be elucidated. Another critical issue is ENCODE also includes several other histone modification marks, DNA methylation, and chromatin accessibility profiling data, providing a rich resource to study murine development (https://www.encodeproject.org/). However, no such completed dataset for mouse ageing exists. As ageing is related to epigenetic alterations (Pal & Tyler, 2016), the generation of large epigenomic datasets consisting of various regulatory factors would greatly benefit ageing research in the future.

5.1.2 How ageing processes influence cancer?

As age is one of the major risk factors for most types of cancer, it is imperative to elucidate how ageing processes could promote cancer progression. Ageing is associated with somatic mutation accumulation and clonal evolution driven by cancer driver genes across most, if not all, human tissues (Kakiuchi & Ogawa, 2021). Thus, this could point toward the earliest event of cancer initiation. Furthermore, a growing body of evidence suggests that tissue microenvironmental changes during ageing could facilitate cancer progression (de Magalhaes, 2013; Fane & Weeraratna, 2020). Cancer evolves through an interplay between somatic mutations and selection in a process that resembles Darwinian evolution (Nowell, 1976; Greaves & Maley, 2012; Podlaha *et al.*, 2012). Results in **Chapter 2** fit well with this concept. I showed that, in general, gene expression changes in ageing and cancer were opposite in most tissues. Ageing is related to the loss of cellular and tissue functions. Thus, the overall functional decline of most cells might create a selective advantage for cells harbouring oncogenic mutations to thrive and develop into cancer. While experiments in blood cancer systems have supported this hypothesis (Henry *et al.*, 2010), evidence in solid tumours is still lacking.

Unlike in most tissues, I found that gene expression changes in ageing and cancer were in a similar direction in the thyroid and uterus. The discrepancy between the transcriptomic

relationship between ageing and cancer in different tissues may indicate that tissue-specific alterations in ageing may affect carcinogenesis and cancer progression differently. In addition, the ageing tissue microenvironment involves alterations in multiple cell types and interactions between them. For example, the accumulation of senescent cells during ageing could help create a pro-tumourigenic environment via releasing pro-inflammatory cytokines as part of SASP (Campisi, 2013). How differences in the ageing microenvironment between tissues contribute to cancer progression remains unknown. Again, a single-cell map of human ageing is expected to shed light on the tissue-specific ageing mechanisms, which would inform how ageing processes influence cancer in a tissue-specific manner. In addition, the advent of tools to analyse intercellular communication and spatial transcriptomic are expected to advance our understanding of the crosstalk between distinct cell types within ageing tissues and reveal potential mechanisms that ageing tissue microenvironment promote tumours (Armingol *et al.*, 2021; Lagger *et al.*, 2021; Liu *et al.*, 2021).

5.1.3 The age-associated cancer molecular landscape: From correlation to causation

In **Chapter 3**, I reported several age-associated genomic, transcriptomic, and DNA methylation patterns across cancers, particularly in gliomas and endometrial cancer. These results and results from other studies clearly show that age does impact the molecular landscape of cancer. These age-related differences may have biological and clinical implications. For example, mutations in *IDH1*, *ATRX*, and *TP53* are more frequently observed in younger glioma patients. The combination of alterations in these three genes blocks NSC differentiation (Modrek *et al.*, 2017). Next, mutations in GATA3, a transcription factor that acts cooperatively with ER and is mutated more frequently in younger breast cancer patients, could promote tumour cell growth and associate with endocrine resistance (Adomas *et al.*, 2014). Furthermore, lower *GATA3* expression is associated with poor prognosis (Mehra *et al.*, 2005). However, mutations in *GATA3* can be both gain-of-function and loss-of-function (Takaku *et al.*, 2018), and it remains unclear whether *GATA3* acts as a tumour suppressor or as an oncogene (Takaku *et al.*, 2015). Therefore, careful investigation of age-related molecular alterations may provide a clue toward age-related mechanisms of cancer development. This knowledge will ultimately help inform cancer treatment strategies.

After identifying numerous age-associated differences in the cancer molecular landscape, the logical next step is understanding why such differences emerge. As discussed in **Chapter 3**,

age-related alterations in the tissue microenvironment might provide different selective advantages for tumours containing distinct molecular alterations, in turn directing tumours to different evolutionary routes. This hypothesis is, however, waiting for experimental evidence. Furthermore, it is also possible that we may still be missing unknown layers of biological and genomic regulation that could be significant in ageing and cancer. Taken together, the findings in **Chapter 3** have generated several important questions for the ageing and cancer communities to investigate in the coming years.

5.1.4 The relationship between development and ageing

The hyperfunction theory of ageing suggests that ageing is partly controlled by processes originating from development (Blagosklonny, 2006; Gems, 2022). However, the molecular links between development and ageing are largely unknown. In **Chapter 4**, I examined the relationship between gene expression and histone modifications (H3K4me3 and H3K27ac) between prenatal development and ageing across four tissues (forebrain, hindbrain, heart, and liver) in mice. In general, findings from the chapter reiterated previous findings that differentially expressed genes in ageing are largely those that were also differentially expressed in development (Somel *et al.*, 2010). In addition, development is related to establishing tissue identity, whereas ageing is associated with losing tissue identity (Izgi *et al.*, 2022). While I revealed a potential contribution of histone modification changes in development and ageing, a major focus in the data analysis in this chapter was at the transcriptomic level. Thus, further studies will be necessary to better elucidate the roles of histone modifications and other epigenetic regulations in regulating both development and ageing.

The identification of molecular changes that are common between development and ageing may point toward developmental pathways that regulate organismal ageing, which will have a wide range of implications, such as identifying new druggable targets for ageing. Our approach to mapping the molecular changes across the lifespan can be complemented by perturbation experiments and pharmacological interventions. For instance, rapamycin treatment early in life has been shown to extend the lifespan of model organisms (Aiello *et al.*, 2022; Shindyapina *et al.*, 2022). While it remains unclear how early life drug treatment could extend lifespan later in life, it has been suggested that epigenetic memory through chromatin modifications and DNA methylation may involve in this process (Lu *et al.*, 2021; Shindyapina *et al.*, 2022). Therefore, detailed mapping of the dynamic epigenomic landscape across the lifespan of an

organism would be necessary to better understand how alterations during development could affect late-life conditions.

5.2 Future directions

As discussed throughout the thesis, works in this study have several limitations. Firstly, all computational genomic analyses in this thesis were done using bulk-level datasets. These bulk-level datasets contain a large sample size and thus can reveal several important patterns regarding the relationship between ageing, cancer, and development. However, the data obtained from these bulk datasets is an average signal from millions of cells and cannot reveal whether the observed patterns originated from alterations in cell proportions or intrinsic changes in gene expression. The continuous development of single-cell assays and a growing number of these single-cell data would allow more in-depth analyses to shed light on the questions that were studied in all three result chapters in this thesis.

Another significant limitation in this thesis is that several findings in this thesis were derived from a correlation. Mechanistic studies are necessary to complement these findings. For example, experimental works are needed to investigate how the ageing tissue microenvironment provides a selective advantage for cancer cells. In addition, while I have identified several age-related molecular patterns across cancers in **Chapter 3**, how these patterns emerge remains to be elucidated. Novel experimental strategies, such as using mouse models of different ages carrying cancer clones with distinct genotypes and advances in single-cell genomics, spatial omics, and statistical methods, would improve our understanding of how age contributes to these emerging age-associated molecular landscapes across cancers. Several questions are expected to be studied in the near future, such as what are the differences in tumour immune landscape according to age, and how do these differences affect response to immunotherapy?

Finally, results in **Chapter 4** suggested potential links between gene expression and epigenetic changes during development and ageing. To study this relationship in more detail, new single-cell multi-omic datasets covering the entire mouse lifespan with a larger sample size are needed. This comprehensive map of dynamic molecular changes across the lifespan and experimental perturbations will provide novel insights into the developmental processes that influence ageing in mammals.

References

Abby, E., Dentro, S. C., Hall, M. W. J., Fowler, J. C., Ong, S. H., Sood, R., ... Jones, P. H. (2021). Notch1 mutation drives clonal expansion in normal esophageal epithelium but impairs tumor growth. *bioRxiv*, 2021.2006.2018.448956.

Adelman, E. R., Huang, H. T., Roisman, A., Olsson, A., Colaprico, A., Qin, T., ... Figueroa, M. E. (2019). Aging Human Hematopoietic Stem Cells Manifest Profound Epigenetic Reprogramming of Enhancers That May Predispose to Leukemia. *Cancer Discov*. **9**, 1080-1101.

Adomas, A. B., Grimm, S. A., Malone, C., Takaku, M., Sims, J. K., & Wade, P. A. (2014). Breast tumor specific mutation in GATA3 affects physiological mechanisms regulating transcription factor turnover. *BMC Cancer*. **14**, 278.

Aiello, G., Sabino, C., Pernici, D., Audano, M., Antonica, F., Gianesello, M., ... Tiberi, L. (2022). Lifespan can be extended during a specific time window early in life. *bioRxiv*, 2022.2002.2018.480980.

Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S., & Stratton, M. R. (2015). Clock-like mutational processes in human somatic cells. *Nat Genet*. **47**, 1402-1407.

Alexandrov, L. B., Ju, Y. S., Haase, K., Van Loo, P., Martincorena, I., Nik-Zainal, S., ... Stratton, M. R. (2016). Mutational signatures associated with tobacco smoking in human cancer. *Science*. **354**, 618-622.

Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat Rev Genet*. **17**, 487-500.

Amemiya, H. M., Kundaje, A., & Boyle, A. P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep*. **9**, 9354.

Anders, C. K., Hsu, D. S., Broadwater, G., Acharya, C. R., Foekens, J. A., Zhang, Y., ... Blackwell, K. L. (2008). Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *J Clin Oncol*. **26**, 3324-3330.

Aramillo Irizar, P., Schauble, S., Esser, D., Groth, M., Frahm, C., Priebe, S., ... Kaleta, C. (2018). Transcriptomic alterations during ageing reflect the shift from cancer to degenerative diseases in the elderly. *Nat Commun*. **9**, 327.

Armingol, E., Officer, A., Harismendy, O., & Lewis, N. E. (2021). Deciphering cell-cell interactions and communication from gene expression. *Nat Rev Genet*. **22**, 71-88.

Ashley, C. W., Da Cruz Paula, A., Kumar, R., Mandelker, D., Pei, X., Riaz, N., ... Weigelt, B. (2019). Analysis of mutational signatures in primary and metastatic endometrial cancer reveals distinct patterns of DNA repair defects and shifts during tumor progression. *Gynecol Oncol*. **152**, 11-19.

Audia, J. E., & Campbell, R. M. (2016). Histone Modifications and Cancer. *Cold Spring Harb Perspect Biol*. **8**, a019521.

Aunan, J. R., Cho, W. C., & Soreide, K. (2017). The Biology of Aging and Cancer: A Brief Overview of Shared and Divergent Molecular Hallmarks. *Aging Dis*. **8**, 628-642.

Avelar, R. A., Ortega, J. G., Tacutu, R., Tyler, E. J., Bennett, D., Binetti, P., ... de Magalhaes, J. P. (2020). A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol*. **21**, 91.

Azim, H. A., Jr., Nguyen, B., Brohee, S., Zoppoli, G., & Sotiriou, C. (2015). Genomic aberrations in young and elderly breast cancer patients. *BMC Med*. **13**, 266.

Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., ... Ding, L. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell*. **173**, 371-385 e318.

Baker, D. J., Childs, B. G., Durik, M., Wijers, M. E., Sieben, C. J., Zhong, J., ... van Deursen, J. M. (2016). Naturally occurring p16(Ink4a)-positive cells shorten healthy lifespan. *Nature*. **530**, 184-189.

Baker, D. J., Wijshake, T., Tchkonia, T., LeBrasseur, N. K., Childs, B. G., van de Sluis, B., ... van Deursen, J. M. (2011). Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. *Nature*. **479**, 232-236.

Barardo, D., Thornton, D., Thoppil, H., Walsh, M., Sharifi, S., Ferreira, S., ... de Magalhaes, J. P. (2017). The DrugAge database of aging-related drugs. *Aging Cell*. **16**, 594-597.

Basisty, N., Kale, A., Jeon, O. H., Kuehnemann, C., Payne, T., Rao, C., ... Schilling, B. (2020). A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLoS Biol*. **18**, e3000599.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*.

Bauer, J., Buttner, P., Murali, R., Okamoto, I., Kolaitis, N. A., Landi, M. T., ... Bastian, B. C. (2011). BRAF mutations in cutaneous melanoma are independently associated with age, anatomic site of the primary tumor, and the degree of solar elastosis at the primary tumor site. *Pigment Cell Melanoma Res*. **24**, 345-351.

Bell, C. G., Lowe, R., Adams, P. D., Baccarelli, A. A., Beck, S., Bell, J. T., ... Rakyan, V. K. (2019). DNA methylation aging clocks: challenges and recommendations. *Genome Biol*. **20**, 249.

Benayoun, B. A., Pollina, E. A., Singh, P. P., Mahmoudi, S., Harel, I., Casey, K. M., ... Brunet, A. (2019). Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. *Genome Res*. **29**, 697-709.

Benisch, P., Schilling, T., Klein-Hitpass, L., Frey, S. P., Seefried, L., Raaijmakers, N., ... Jakob, F. (2012). The transcriptional profile of mesenchymal stem cell populations in primary osteoporosis is distinct and shows overexpression of osteogenic inhibitors. *PLoS One*. **7**, e45142.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple

Testing. *J. R. Statist. Soc. B*. **57**, 289-300.

Benz, C. C. (2008). Impact of aging on the biology of breast cancer. *Crit Rev Oncol Hematol*. **66**, 65-74.

Berben, L., Floris, G., Wildiers, H., & Hatse, S. (2021). Cancer and Aging: Two Tightly Interconnected Biological Processes. *Cancers (Basel)*. **13**.

Berg, M., Danielsen, S. A., Ahlquist, T., Merok, M. A., Agesen, T. H., Vatn, M. H., ... Lothe, R. A. (2010). DNA sequence profiles of the colorectal cancer critical gene set KRAS-BRAF-PIK3CA-PTEN-TP53 related to age at disease onset. *PLoS One*. **5**, e13978.

Berger, A. C., Korkut, A., Kanchi, R. S., Hegde, A. M., Lenoir, W., Liu, W., ... Akbani, R. (2018). A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. *Cancer Cell*. **33**, 690-705 e699.

Bernardes de Jesus, B., Schneeberger, K., Vera, E., Tejera, A., Harley, C. B., & Blasco, M. A. (2011). The telomerase activator TA-65 elongates short telomeres and increases health span of adult/old mice without increasing cancer incidence. *Aging Cell*. **10**, 604-621.

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., ... Meyerson, M. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*. **463**, 899-905.

Bhatia, B., Jiang, M., Suraneni, M., Patrawala, L., Badeaux, M., Schneider-Broussard, R., ... Tang, D. G. (2008). Critical and distinct roles of p16 and telomerase in regulating the proliferative life span of normal human prostate epithelial progenitor cells. *J Biol Chem*. **283**, 27957-27972.

Bick, A. G., Weinstock, J. S., Nandakumar, S. K., Fulco, C. P., Bao, E. L., Zekavat, S. M., ... Natarajan, P. (2020). Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature*. **586**, 763-768.

Bielski, C. M., Zehir, A., Penson, A. V., Donoghue, M. T. A., Chatila, W., Armenia, J., ... Taylor, B. S. (2018). Genome doubling shapes the evolution and prognosis of advanced cancers. *Nat Genet*. **50**, 1189-1195.

Binet, R., Ythier, D., Robles, A. I., Collado, M., Larrieu, D., Fonti, C., ... Pedeux, R. (2009). WNT16B is a new marker of cellular senescence that regulates p53 activity and the phosphoinositide 3-kinase/AKT pathway. *Cancer Res*. **69**, 9183-9191.

Blagosklonny, M. V. (2006). Aging and immortality: quasi-programmed senescence and its pharmacologic inhibition. *Cell Cycle*. **5**, 2087-2102.

Blokzijl, F., de Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., ... van Boxtel, R. (2016). Tissue-specific mutation accumulation in human adult stem cells during life. *Nature*. **538**, 260-264.

Bonneville, R., Krook, M. A., Kautto, E. A., Miya, J., Wing, M. R., Chen, H. Z., ... Roychowdhury, S. (2017). Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precis Oncol*. **2017**.

Borghesan, M., Fafian-Labora, J., Eleftheriadou, O., Carpintero-Fernandez, P., Paez-Ribes, M., Vizcay-Barrena, G., ... O'Loghlen, A. (2019). Small Extracellular Vesicles Are Key Regulators of Non-cell Autonomous Intercellular Communication in Senescence via the Interferon Protein IFITM3. *Cell Rep*. **27**, 3956-3971 e3956.

Bozdag, S., Li, A., Riddick, G., Kotliarov, Y., Baysan, M., Iwamoto, F. M., ... Fine, H. A. (2013). Age-specific signatures of glioblastoma at the genomic, genetic, and epigenetic levels. *PLoS One*. **8**, e62982.

Brack, A. S., Conboy, M. J., Roy, S., Lee, M., Kuo, C. J., Keller, C., & Rando, T. A. (2007). Increased Wnt signaling during aging alters muscle stem cell fate and increases fibrosis. *Science*. **317**, 807-810.

Brennan, C. W., Verhaak, R. G., McKenna, A., Campos, B., Noushmehr, H., Salama, S. R., ... Network, T. R. (2013). The somatic genomic landscape of glioblastoma. *Cell*. **155**, 462-477.

Brosh, R., Sarig, R., Natan, E. B., Molchadsky, A., Madar, S., Bornstein, C., ... Rotter, V. (2010). p53-dependent transcriptional regulation of EDA2R and its involvement in chemotherapy-induced hair loss. *FEBS Lett*. **584**, 2473-2477.

Brunner, S. F., Roberts, N. D., Wylie, L. A., Moore, L., Aitken, S. J., Davies, S. E., ... Campbell, P. J. (2019). Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature*. **574**, 538-542.

Campbell, B. B., Light, N., Fabrizio, D., Zatzman, M., Fuligni, F., de Borja, R., ... Shlien, A. (2017). Comprehensive Analysis of Hypermutation in Human Cancer. *Cell*. **171**, 1042-1056 e1010.

Campisi, J. (2013). Aging, cellular senescence, and cancer. *Annu Rev Physiol*. **75**, 685-705.

Campisi, J., & Yaswen, P. (2009). Aging and cancer cell biology, 2009. *Aging Cell*. **8**, 221-225.

Cancer Genome Atlas, N. (2012a). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. **487**, 330-337.

Cancer Genome Atlas, N. (2012b). Comprehensive molecular portraits of human breast tumours. *Nature*. **490**, 61-70.

Cancer Genome Atlas Research, N. (2014a). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. **513**, 202-209.
Cancer Genome Atlas Research, N. (2014b). Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. **511**, 543-550.
Cancer Genome Atlas Research, N. (2014c). Integrated genomic characterization of papillary thyroid carcinoma. *Cell*. **159**, 676-690.
Cancer Genome Atlas Research, N. (2015). The Molecular Taxonomy of Primary Prostate Cancer. *Cell*. **163**, 1011-1025.
Cancer Genome Atlas Research, N., Analysis Working Group: Asan, U., Agency, B. C. C., Brigham, Women's, H., Broad, I., ... Project Team: National Institutes of, H. (2017). Integrated genomic characterization of oesophageal carcinoma. *Nature*. **541**, 169-175.
Cancer Genome Atlas Research, N., Brat, D. J., Verhaak, R. G., Aldape, K. D., Yung, W. K., Salama, S. R., ... Zhang, J. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med*. **372**, 2481-2498.
Cancer Genome Atlas Research, N., Kandoth, C., Schultz, N., Cherniack, A. D., Akbani, R., Liu, Y., ... Levine, D. A. (2013a). Integrated genomic characterization of endometrial carcinoma. *Nature*. **497**, 67-73.
Cancer Genome Atlas Research, N., Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R., Ozenberger, B. A., ... Stuart, J. M. (2013b). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. **45**, 1113-1120.
Cancer Genome Atlas Research Network. Electronic address, w. b. e., & Cancer Genome Atlas Research, N. (2017). Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell*. **169**, 1327-1341 e1323.
Cao, K., Blair, C. D., Faddah, D. A., Kieckhaefer, J. E., Olive, M., Erdos, M. R., ... Collins, F. S. (2011). Progerin and telomere dysfunction collaborate to trigger cellular senescence in normal human fibroblasts. *J Clin Invest*. **121**, 2833-2844.
Ceccarelli, M., Barthel, F. P., Malta, T. M., Sabedot, T. S., Salama, S. R., Murray, B. A., ... Verhaak, R. G. (2016). Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell*. **164**, 550-563.
Chalmers, Z. R., Connelly, C. F., Fabrizio, D., Gay, L., Ali, S. M., Ennis, R., ... Frampton, G. M. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med*. **9**, 34.
Chang, J., Wang, Y., Shao, L., Laberge, R. M., Demaria, M., Campisi, J., ... Zhou, D. (2016). Clearance of senescent cells by ABT263 rejuvenates aged hematopoietic stem cells in mice. *Nat Med*. **22**, 78-83.
Charles Joly Beauparlant, A. L., Eric Fournier, and Arnaud Droit (2020). ENCODExplorer: A compilation of ENCODE metadataed^eds).
Chatsirisupachai, K., Lesluyes, T., Paraoan, L., Van Loo, P., & de Magalhaes, J. P. (2021). An integrative analysis of the age-associated multi-omic landscape across cancers. *Nat Commun*. **12**, 2345.
Chatsirisupachai, K., Palmer, D., Ferreira, S., & de Magalhaes, J. P. (2019). A human tissue-specific transcriptomic analysis reveals a complex relationship between aging, cancer, and cellular senescence. *Aging Cell*. **18**, e13041.
Chen, H., & Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics*. **12**, 35.
Chen, H., Iglesias, M. A., Caruso, V., & Morris, M. J. (2011). Maternal cigarette smoke exposure contributes to glucose intolerance and decreased brain insulin action in mice offspring independent of maternal diet. *PLoS One*. **6**, e27260.

Chen, J. H., Martin-Gronert, M. S., Tarry-Adkins, J., & Ozanne, S. E. (2009). Maternal protein restriction affects postnatal growth and the expression of key proteins involved in lifespan regulation in mice. *PLoS One*. **4**, e4950.

Childs, B. G., Durik, M., Baker, D. J., & van Deursen, J. M. (2015). Cellular senescence in aging and age-related disease: from mechanisms to therapy. *Nat Med*. **21**, 1424-1435.

Cieslik, M., & Chinnaiyan, A. M. (2018). Cancer transcriptome profiling at the juncture of clinical translation. *Nat Rev Genet*. **19**, 93-109.

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., ... Noushmehr, H. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. **44**, e71.

Cole, J. J., Robertson, N. A., Rather, M. I., Thomson, J. P., McBryan, T., Sproul, D., ... Adams, P. D. (2017). Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol*. **18**, 58.

Collado-Torres, L., Nellore, A., Kammers, K., Ellis, S. E., Taub, M. A., Hansen, K. D., ... Leek, J. T. (2017). Reproducible RNA-seq analysis using recount2. *Nat Biotechnol*. **35**, 319-321.

Collisson, E. A., Cho, R. J., & Gray, J. W. (2012). What are we learning from the cancer genome? *Nat Rev Clin Oncol*. **9**, 621-630.

Colom, B., Alcolea, M. P., Piedrafita, G., Hall, M. W. J., Wabik, A., Dentro, S. C., ... Jones, P. H. (2020). Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. *Nat Genet*. **52**, 604-614.

Colom, B., Herms, A., Hall, M. W. J., Dentro, S. C., King, C., Sood, R. K., ... Jones, P. H. (2021). Mutant clones in normal epithelium outcompete and eliminate emerging tumours. *Nature*. **598**, 510-514.

Consortium, A. P. G. (2017). AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov*. **7**, 818-831.

Consortium, E. P., Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shoresh, N., ... Weng, Z. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. **583**, 699-710.

Consortium, G. T. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. **45**, 580-585.

Consortium, G. T. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. **348**, 648-660.

Consortium, I. T. P.-C. A. o. W. G. (2020). Pan-cancer analysis of whole genomes. *Nature*. **578**, 82-93.

Coppe, J. P., Desprez, P. Y., Krtolica, A., & Campisi, J. (2010). The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev Pathol*. **5**, 99-118.

Coppe, J. P., Kauser, K., Campisi, J., & Beausejour, C. M. (2006). Secretion of vascular endothelial growth factor by primary human fibroblasts at senescence. *J Biol Chem*. **281**, 29568-29574.

Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., ... Aparicio, S. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. **486**, 346-352.

Damle, R. P., Dravid, N. V., Suryawanshi, K. H., Gadre, A. S., Bagale, P. S., & Ahire, N. (2013). Clinicopathological Spectrum of Endometrial Changes in Peri-menopausal and Post-menopausal Abnormal Uterine Bleeding: A 2 Years Study. *J Clin Diagn Res*. **7**, 2774-2776.

Davoli, T., Uno, H., Wooten, E. C., & Elledge, S. J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*. **355**.

de Magalhaes, J. P. (2012). Programmatic features of aging originating in development: aging mechanisms beyond molecular damage? *FASEB J*. **26**, 4821-4826.
de Magalhaes, J. P. (2013). How ageing processes influence cancer. *Nat Rev Cancer*. **13**, 357-365.
de Magalhaes, J. P., & Church, G. M. (2005). Genomes optimize reproduction: aging as a consequence of the developmental program. *Physiology (Bethesda)*. **20**, 252-259.
de Magalhaes, J. P., Curado, J., & Church, G. M. (2009). Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics*. **25**, 875-881.
de Magalhães, J. P., & Stuart-Hamilton, I. (2011). The biology of ageing: a primer. In *An Introduction to Gerontology*). Cambridge, UK: Cambridge University Press, pp. 21-47.
de Vries, M., Faiz, A., Woldhuis, R. R., Postma, D. S., de Jong, T. V., Sin, D. D., ... Brandsma, C. A. (2017). Lung tissue gene-expression signature for the ageing lung in COPD. *Thorax*.
Demanelis, K., Jasmine, F., Chen, L. S., Chernoff, M., Tong, L., Delgado, D., ... Pierce, B. L. (2020). Determinants of telomere length across human tissues. *Science*. **369**.
Demaria, M., Ohtani, N., Youssef, S. A., Rodier, F., Toussaint, W., Mitchell, J. R., ... Campisi, J. (2014). An essential role for senescent cells in optimal wound healing through secretion of PDGF-AA. *Dev Cell*. **31**, 722-733.
Desai, P., Mencia-Trinchant, N., Savenkov, O., Simon, M. S., Cheang, G., Lee, S., ... Hassane, D. C. (2018). Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat Med*. **24**, 1015-1023.
Dewhurst, S. M., McGranahan, N., Burrell, R. A., Rowan, A. J., Gronroos, E., Endesfelder, D., ... Swanton, C. (2014). Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. *Cancer Discov*. **4**, 175-185.
Di Micco, R., Krizhanovsky, V., Baker, D., & d'Adda di Fagagna, F. (2021). Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nat Rev Mol Cell Biol*. **22**, 75-95.
Ding, Y., Wu, H., Warden, C., Steele, L., Liu, X., Iterson, M. V., ... Neuhausen, S. L. (2016). Gene Expression Differences in Prostate Cancers between Young and Old Men. *PLoS Genet*. **12**, e1006477.
Djiogue, S., Nwabo Kamdje, A. H., Vecchio, L., Kipanyula, M. J., Farahna, M., Aldebasi, Y., & Seke Etet, P. F. (2013). Insulin resistance and cancer: the role of insulin and IGFs. *Endocr Relat Cancer*. **20**, R1-R17.
Dmitrijeva, M., Ossowski, S., Serrano, L., & Schaefer, M. H. (2018). Tissue-specific DNA methylation loss during ageing and carcinogenesis is linked to chromosome structure, replication timing and cell division rates. *Nucleic Acids Res*. **46**, 7022-7039.
Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. **29**, 15-21.
Donertas, H. M., Izgi, H., Kamacioglu, A., He, Z., Khaitovich, P., & Somel, M. (2017). Gene expression reversal toward pre-adult levels in the aging human brain and age-related loss of cellular identity. *Sci Rep*. **7**, 5894.
Dubrovska, A., Kim, S., Salamone, R. J., Walker, J. R., Maira, S. M., Garcia-Echeverria, C., ... Reddy, V. A. (2009). The role of PTEN/Akt/PI3K signaling in the maintenance and viability of prostate cancer stem-like cell populations. *Proc Natl Acad Sci U S A*. **106**, 268-273.
Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*. **4**, 1184-1191.

Ellrott, K., Bailey, M. H., Saksena, G., Covington, K. R., Kandoth, C., Stewart, C., ... Cancer Genome Atlas Research, N. (2018). Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. *Cell Syst*. **6**, 271-281 e277.

Enge, M., Arda, H. E., Mignardi, M., Beausang, J., Bottino, R., Kim, S. K., & Quake, S. R. (2017). Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell*. **171**, 321-330 e314.

Erbe, R., Wang, Z., Wu, S., Xiu, J., Zaidi, N., La, J., ... Fertig, E. J. (2021). Evaluating the impact of age on immune checkpoint therapy biomarkers. *Cell Rep*. **36**, 109599.

Erbe, R., Wang, Z., Zaidi, N., Topper, M., Baylin, S., Jaffee, E. M., ... Fertig, E. J. (2020). Aging interacts with tumor biology to produce major changes in the immune tumor microenvironment. *bioRxiv*.

Faget, D. V., Ren, Q., & Stewart, S. A. (2019). Unmasking senescence: context-dependent effects of SASP in cancer. *Nat Rev Cancer*. **19**, 439-453.

Fane, M., & Weeraratna, A. T. (2020). How the ageing microenvironment influences tumour progression. *Nat Rev Cancer*. **20**, 89-106.

Feinberg, A. P., Koldobskiy, M. A., & Gondor, A. (2016). Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat Rev Genet*. **17**, 284-299.

Fernandez-Antoran, D., Piedrafita, G., Murai, K., Ong, S. H., Herms, A., Frezza, C., & Jones, P. H. (2019). Outcompeting p53-Mutant Cells in the Normal Esophagus by Redox Manipulation. *Cell Stem Cell*. **25**, 329-341 e326.

Ferreira, P. G., Munoz-Aguirre, M., Reverter, F., Sa Godinho, C. P., Sousa, A., Amadoz, A., ... Guigo, R. (2018). The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nat Commun*. **9**, 490.

Ferrucci, L., & Fabbri, E. (2018). Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. *Nat Rev Cardiol*. **15**, 505-522.

Feulner, L., Najafabadi, H. S., Tanguay, S., Rak, J., & Riazalhosseini, Y. (2019). Age-related variations in gene expression patterns of renal cell carcinoma. *Urol Oncol*. **37**, 166-175.

Fontana, L., Partridge, L., & Longo, V. D. (2010). Extending healthy life span--from yeast to humans. *Science*. **328**, 321-326.

Franceschi, C., Garagnani, P., Morsiani, C., Conte, M., Santoro, A., Grignolio, A., ... Salvioli, S. (2018a). The Continuum of Aging and Age-Related Diseases: Common Mechanisms but Different Rates. *Front Med (Lausanne)*. **5**, 61.

Franceschi, C., Garagnani, P., Parini, P., Giuliani, C., & Santoro, A. (2018b). Inflammaging: a new immune-metabolic viewpoint for age-related diseases. *Nat Rev Endocrinol*. **14**, 576-590.

Franco, I., Johansson, A., Olsson, K., Vrtacnik, P., Lundin, P., Helgadottir, H. T., ... Eriksson, M. (2018). Somatic mutagenesis in satellite cells associates with human skeletal muscle aging. *Nat Commun*. **9**, 800.

Frobel, J., Hemeda, H., Lenz, M., Abagnale, G., Joussen, S., Denecke, B., ... Wagner, W. (2014). Epigenetic rejuvenation of mesenchymal stromal cells derived from induced pluripotent stem cells. *Stem Cell Reports*. **3**, 414-422.

Fulop, T., Kotb, R., Fortin, C. F., Pawelec, G., de Angelis, F., & Larbi, A. (2010). Potential role of immunosenescence in cancer development. *Ann N Y Acad Sci*. **1197**, 158-165.

Fulop, T., Larbi, A., Dupuis, G., Le Page, A., Frost, E. H., Cohen, A. A., ... Franceschi, C. (2017). Immunosenescence and Inflamm-Aging As Two Sides of the Same Coin: Friends or Foes? *Front Immunol*. **8**, 1960.

Garcia-Nieto, P. E., Morrison, A. J., & Fraser, H. B. (2019). The somatic mutation landscape of the human body. *Genome Biol*. **20**, 298.

Garraway, L. A., & Lander, E. S. (2013). Lessons from the cancer genome. *Cell*. **153**, 17-37.

Gasek, N. S., Kuchel, G. A., Kirkland, J. L., & Xu, M. (2021). Strategies for Targeting Senescent Cells in Human Disease. *Nat Aging*. **1**, 870-879.
Gemechu, J. M., & Bentivoglio, M. (2012). T Cell Recruitment in the Brain during Normal Aging. *Front Cell Neurosci*. **6**, 38.
Gems, D. (2022). The hyperfunction theory: An emerging paradigm for the biology of aging. *Ageing Res Rev*. **74**, 101557.
Gems, D., & de Magalhaes, J. P. (2021). The hoverfly and the wasp: A critique of the hallmarks of aging as a paradigm. *Ageing Res Rev*. **70**, 101407.
Genovese, G., Kahler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., ... McCarroll, S. A. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*. **371**, 2477-2487.
Gerhauser, C., Favero, F., Risch, T., Simon, R., Feuerbach, L., Assenov, Y., ... Weischenfeldt, J. (2018). Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. *Cancer Cell*. **34**, 996-1011 e1018.
Glass, D., Vinuela, A., Davies, M. N., Ramasamy, A., Parts, L., Knowles, D., ... Spector, T. D. (2013). Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol*. **14**, R75.
Gnani, D., Crippa, S., Della Volpe, L., Rossella, V., Conti, A., Lettera, E., ... Di Micco, R. (2019). An early-senescence state in aged mesenchymal stromal cells contributes to hematopoietic stem and progenitor cell clonogenic impairment through the activation of a pro-inflammatory program. *Aging Cell*. **18**, e12933.
Gonzalez-Meljem, J. M., Apps, J. R., Fraser, H. C., & Martinez-Barbera, J. P. (2018). Paracrine roles of cellular senescence in promoting tumourigenesis. *Br J Cancer*. **118**, 1283-1288.
Gonzalez-Suarez, E., Samper, E., Ramirez, A., Flores, J. M., Martin-Caballero, J., Jorcano, J. L., & Blasco, M. A. (2001). Increased epidermal tumors and increased skin wound healing in transgenic mice overexpressing the catalytic subunit of telomerase, mTERT, in basal keratinocytes. *EMBO J*. **20**, 2619-2630.
Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. **17**, 333-351.
Gorgoulis, V., Adams, P. D., Alimonti, A., Bennett, D. C., Bischof, O., Bishop, C., ... Demaria, M. (2019). Cellular Senescence: Defining a Path Forward. *Cell*. **179**, 813-827.
Gorkin, D. U., Barozzi, I., Zhao, Y., Zhang, Y., Huang, H., Lee, A. Y., ... Ren, B. (2020). An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature*. **583**, 744-751.
Greaves, M., & Maley, C. C. (2012). Clonal evolution in cancer. *Nature*. **481**, 306-313.
Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: a problem hiding in plain sight. *BMJ*. **352**, i1981.
Grobner, S. N., Worst, B. C., Weischenfeldt, J., Buchhalter, I., Kleinheinz, K., Rudneva, V. A., ... Pfister, S. M. (2018). The landscape of genomic alterations across childhood cancers. *Nature*. **555**, 321-327.
Grossmann, S., Hooks, Y., Wilson, L., Moore, L., O'Neill, L., Martincorena, I., ... Campbell, P. J. (2021). Development, maturation, and maintenance of human prostate inferred from somatic mutations. *Cell Stem Cell*. **28**, 1262-1274 e1265.
Gruber, H. E., Hoelscher, G. L., Ingram, J. A., Zinchenko, N., & Hanley, E. N., Jr. (2010). Senescent vs. non-senescent cells in the human annulus in vivo: cell harvest with laser capture microdissection and gene expression studies with microarray analysis. *BMC Biotechnol*. **10**, 5.
Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. **32**, 2847-2849.

Guerrero, A., Iglesias, C., Raguz, S., Floridia, E., Gil, J., Pombo, C. M., & Zalvide, J. (2015). The cerebral cavernous malformation 3 gene is necessary for senescence induction. *Aging Cell*. **14**, 274-283.

Guney, I., Wu, S., & Sedivy, J. M. (2006). Reduced c-Myc signaling triggers telomere-independent senescence by regulating Bmi-1 and p16(INK4a). *Proc Natl Acad Sci U S A*. **103**, 3645-3650.

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discov*. **12**, 31-46.

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*. **100**, 57-70.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*. **144**, 646-674.

Hanson, M. A., Cooper, C., Aihie Sayer, A., Eendebak, R. J., Clough, G. F., & Beard, J. R. (2016). Developmental aspects of a life course approach to healthy ageing. *J Physiol*. **594**, 2147-2160.

Harbers, L., Agostini, F., Nicos, M., Poddighe, D., Bienko, M., & Crosetto, N. (2021). Somatic Copy Number Alterations in Human Cancers: An Analysis of Publicly Available Data From The Cancer Genome Atlas. *Front Oncol*. **11**, 700568.

Hardy, K., Mansfield, L., Mackay, A., Benvenuti, S., Ismail, S., Arora, P., ... Jat, P. S. (2005). Transcriptional networks and cellular senescence in human mammary fibroblasts. *Mol Biol Cell*. **16**, 943-953.

Harrison, D. E., Strong, R., Sharp, Z. D., Nelson, J. F., Astle, C. M., Flurkey, K., ... Miller, R. A. (2009). Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature*. **460**, 392-395.

Hartman, Z. C., Poage, G. M., den Hollander, P., Tsimelzon, A., Hill, J., Panupinthu, N., ... Brown, P. H. (2013). Growth of triple-negative breast cancer cells relies upon coordinate autocrine expression of the proinflammatory cytokines IL-6 and IL-8. *Cancer Res*. **73**, 3470-3480.

Hayflick, L., & Moorhead, P. S. (1961). The serial cultivation of human diploid cell strains. *Exp Cell Res*. **25**, 585-621.

Heinze, G., & Ploner, M. (2018). logistf: Firth's Bias-Reduced Logistic Regression.

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Stat Med*. **21**, 2409-2419.

Henry, C. J., Casas-Selves, M., Kim, J., Zaberezhnyy, V., Aghili, L., Daniel, A. E., ... DeGregori, J. (2015). Aging-associated inflammation promotes selection for adaptive oncogenic events in B cell progenitors. *J Clin Invest*. **125**, 4666-4680.

Henry, C. J., Marusyk, A., Zaberezhnyy, V., Adane, B., & DeGregori, J. (2010). Declining lymphoid progenitor fitness promotes aging-associated leukemogenesis. *Proc Natl Acad Sci U S A*. **107**, 21713-21718.

Herbig, U., Ferreira, M., Condel, L., Carey, D., & Sedivy, J. M. (2006). Cellular senescence in aging primates. *Science*. **311**, 1257.

Hernandez-Segura, A., de Jong, T. V., Melov, S., Guryev, V., Campisi, J., & Demaria, M. (2017). Unmasking Transcriptional Heterogeneity in Senescent Cells. *Curr Biol*. **27**, 2652-2660 e2654.

Hernandez-Segura, A., Nehme, J., & Demaria, M. (2018). Hallmarks of Cellular Senescence. *Trends Cell Biol*. **28**, 436-453.

Hernando, B., Dietzen, M., Parra, G., Gil-Barrachina, M., Pitarch, G., Mahiques, L., ... Martinez-Cadenas, C. (2021). The effect of age on the acquisition and selection of cancer driver mutations in sun-exposed normal skin. *Ann Oncol*. **32**, 412-421.

Hernando-Herraez, I., Evano, B., Stubbs, T., Commere, P. H., Jan Bonder, M., Clark, S., ... Reik, W. (2019). Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nat Commun*. **10**, 4361.

Hieronymus, H., Murali, R., Tin, A., Yadav, K., Abida, W., Moller, H., ... Sawyers, C. L. (2018). Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. *Elife*. **7**.

Higa, K. C., & DeGregori, J. (2019). Decoy fitness peaks, tumor suppression, and aging. *Aging Cell*. **18**, e12938.

Hua, H., Kong, Q., Zhang, H., Wang, J., Luo, T., & Jiang, Y. (2019). Targeting mTOR for cancer therapy. *J Hematol Oncol*. **12**, 71.

Huang, K. L., Mashl, R. J., Wu, Y., Ritter, D. I., Wang, J., Oh, C., ... Ding, L. (2018). Pathogenic Germline Variants in 10,389 Adult Cancers. *Cell*. **173**, 355-370 e314.

Huang, Z., Chen, B., Liu, X., Li, H., Xie, L., Gao, Y., ... Su, W. (2021). Effects of sex and aging on the immune cell landscape as assessed by single-cell transcriptomic analysis. *Proc Natl Acad Sci U S A*. **118**.

Idda, M. L., McClusky, W. G., Lodde, V., Munk, R., Abdelmohsen, K., Rossi, M., & Gorospe, M. (2020). Survey of senescent cell markers with age in human tissues. *Aging (Albany NY)*. **12**, 4052-4066.

Imai, Y., Takahashi, A., Hanyu, A., Hori, S., Sato, S., Naka, K., ... Hara, E. (2014). Crosstalk between the Rb pathway and AKT signaling forms a quiescence-senescence switch. *Cell Rep*. **7**, 194-207.

International Cancer Genome, C., Hudson, T. J., Anderson, W., Artez, A., Barker, A. D., Bell, C., ... Yang, H. (2010). International network of cancer genome projects. *Nature*. **464**, 993-998.

Izgi, H., Han, D., Isildak, U., Huang, S., Kocabiyik, E., Khaitovich, P., ... Donertas, H. M. (2022). Inter-tissue convergence of gene expression during ageing suggests age-related loss of tissue and cellular identity. *Elife*. **11**.

Jacobs, K. B., Yeager, M., Zhou, W., Wacholder, S., Wang, Z., Rodriguez-Santiago, B., ... Chanock, S. J. (2012). Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet*. **44**, 651-658.

Jaiswal, S., & Ebert, B. L. (2019). Clonal hematopoiesis in human aging and disease. *Science*. **366**.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B. G., ... Ebert, B. L. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. **371**, 2488-2498.

Jaiswal, S., Natarajan, P., Silver, A. J., Gibson, C. J., Bick, A. G., Shvartz, E., ... Ebert, B. L. (2017). Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med*. **377**, 111-121.

Janes, J., Dong, Y., Schoof, M., Serizay, J., Appert, A., Cerrato, C., ... Ahringer, J. (2018). Chromatin accessibility dynamics across C. elegans development and ageing. *Elife*. **7**.

Jeon, O. H., Kim, C., Laberge, R. M., Demaria, M., Rathod, S., Vasserot, A. P., ... Elisseeff, J. H. (2017). Local clearance of senescent cells attenuates the development of post-traumatic osteoarthritis and creates a pro-regenerative environment. *Nat Med*. **23**, 775-781.

Jeong, I., Lim, J. H., Park, J. S., & Oh, Y. M. (2020). Aging-related changes in the gene expression profile of human lungs. *Aging (Albany NY)*. **12**, 21391-21403.

Jezequel, P., Sharif, Z., Lasla, H., Gouraud, W., Guerin-Charbonnel, C., Campion, L., ... Campone, M. (2015). Gene-expression signature functional annotation of breast cancer tumours in function of age. *BMC Med Genomics*. **8**, 80.

Jin, K. (2010). Modern Biological Theories of Aging. *Aging Dis*. **1**, 72-74.

Johnson, A. A., Akman, K., Calimport, S. R., Wuttke, D., Stolzing, A., & de Magalhaes, J. P. (2012). The role of DNA methylation in aging, rejuvenation, and age-related disease. *Rejuvenation Res*. **15**, 483-494.

Johung, K., Goodwin, E. C., & DiMaio, D. (2007). Human papillomavirus E7 repression in cervical carcinoma cells initiates a transcriptional cascade driven by the retinoblastoma family, resulting in senescence. *J Virol*. **81**, 2102-2116.

Jong, H. L., Mustafa, M. R., Vanhoutte, P. M., AbuBakar, S., & Wong, P. F. (2013). MicroRNA 299-3p modulates replicative senescence in endothelial cells. *Physiol Genomics*. **45**, 256-267.

Kakiuchi, N., & Ogawa, S. (2021). Clonal expansion in non-cancer tissues. *Nat Rev Cancer*. **21**, 239-256.

Kan, Z., Ding, Y., Kim, J., Jung, H. H., Chung, W., Lal, S., ... Park, Y. H. (2018). Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. *Nat Commun*. **9**, 1725.

Karlic, R., Chung, H. R., Lasserre, J., Vlahovicek, K., & Vingron, M. (2010). Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A*. **107**, 2926-2931.

Kassambara, A. (2020a). ggpubr: 'ggplot2' Based Publication Ready Plots.

Kassambara, A. M., F. (2020b). factoextra: Extract and Visualize the Results of Multivariate Data Analysesed^eds).

Kaur, A., Webster, M. R., Marchbank, K., Behera, R., Ndoye, A., Kugel, C. H., 3rd, ... Weeraratna, A. T. (2016). sFRP2 in the aged microenvironment drives melanoma metastasis and therapy resistance. *Nature*. **532**, 250-254.

Kennedy, B. K., Berger, S. L., Brunet, A., Campisi, J., Cuervo, A. M., Epel, E. S., ... Sierra, F. (2014). Geroscience: linking aging to chronic disease. *Cell*. **159**, 709-713.

Kim, B. C., Jeong, H. O., Park, D., Kim, C. H., Lee, E. K., Kim, D. H., ... Chung, H. Y. (2015). Profiling age-related epigenetic markers of stomach adenocarcinoma in young and old subjects. *Cancer Inform*. **14**, 47-54.

Kim, T. M., Laird, P. W., & Park, P. J. (2013a). The landscape of microsatellite instability in colorectal and endometrial cancer genomes. *Cell*. **155**, 858-868.

Kim, Y. H., Choi, Y. W., Lee, J., Soh, E. Y., Kim, J. H., & Park, T. J. (2017). Senescent tumor cells lead the collective invasion in thyroid cancer. *Nat Commun*. **8**, 15208.

Kim, Y. M., Byun, H. O., Jee, B. A., Cho, H., Seo, Y. H., Kim, Y. S., ... Yoon, G. (2013b). Implications of time-series gene expression profiles of replicative senescence. *Aging Cell*. **12**, 622-634.

Kirkwood, T. B. (2005). Understanding the odd science of aging. *Cell*. **120**, 437-447.

Korber, V., Yang, J., Barah, P., Wu, Y., Stichel, D., Gu, Z., ... Lichter, P. (2019). Evolutionary Trajectories of IDH(WT) Glioblastomas Reveal a Common Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis. *Cancer Cell*. **35**, 692-704 e612.

Krtolica, A., Parrinello, S., Lockett, S., Desprez, P. Y., & Campisi, J. (2001). Senescent fibroblasts promote epithelial cell growth and tumorigenesis: a link between cancer and aging. *Proc Natl Acad Sci U S A*. **98**, 12072-12077.

La Manno, G., Siletti, K., Furlan, A., Gyllborg, D., Vinsland, E., Mossi Albiach, A., ... Linnarsson, S. (2021). Molecular architecture of the developing mouse brain. *Nature*. **596**, 92-96.

Laberge, R. M., Awad, P., Campisi, J., & Desprez, P. Y. (2012). Epithelial-mesenchymal transition induced by senescent fibroblasts. *Cancer Microenviron*. **5**, 39-44.

Laberge, R. M., Sun, Y., Orjalo, A. V., Patil, C. K., Freund, A., Zhou, L., ... Campisi, J. (2015). MTOR regulates the pro-tumorigenic senescence-associated secretory phenotype by promoting IL1A translation. *Nat Cell Biol*. **17**, 1049-1061.

Laconi, E., Marongiu, F., & DeGregori, J. (2020). Cancer as a disease of old age: changing mutational and microenvironmental landscapes. *Br J Cancer*. **122**, 943-952.

Lagger, C., Ursu, E., Equey, A., Avelar, R. A., Pisco, A. O., Tacutu, R., & de Magalhães, J. P. (2021). scAgeCom: a murine atlas of age-related changes in intercellular communication inferred with the package scDiffCom. *bioRxiv*, 2021.2008.2013.456238.

Lai, M. C., Ruiz-Velasco, M., Arnold, C., Sigalova, O., Bunina, D., Berest, I., ... Zaugg, J. B. (2021). Enhancer-priming in ageing human bone marrow mesenchymal stromal cells contributes to immune traits. *bioRxiv*, 2021.2009.2003.458728.

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., ... Snyder, M. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. **22**, 1813-1831.

Laurie, C. C., Laurie, C. A., Rice, K., Doheny, K. F., Zelnick, L. R., McHugh, C. P., ... Weir, B. S. (2012). Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet*. **44**, 642-650.

Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., ... Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. **505**, 495-501.

Lawson, A. R. J., Abascal, F., Coorens, T. H. H., Hooks, Y., O'Neill, L., Latimer, C., ... Martincorena, I. (2020). Extensive heterogeneity in somatic mutation and selection in the human bladder. *Science*. **370**, 75-82.

Lederle, W., Depner, S., Schnur, S., Obermueller, E., Catone, N., Just, A., ... Mueller, M. M. (2011). IL-6 promotes malignant growth of skin SCCs by regulating a network of autocrine and paracrine cytokines. *Int J Cancer*. **128**, 2803-2814.

Lee, W., Wang, Z., Saffern, M., Jun, T., & Huang, K. L. (2021). Genomic and molecular features distinguish young adult cancer from later-onset cancer. *Cell Rep*. **37**, 110005.

Lee-Six, H., Olafsson, S., Ellis, P., Osborne, R. J., Sanders, M. A., Moore, L., ... Stratton, M. R. (2019). The landscape of somatic mutation in normal colorectal epithelial cells. *Nature*. **574**, 532-537.

Leonardi, G. C., Accardi, G., Monastero, R., Nicoletti, F., & Libra, M. (2018). Ageing: from inflammation to cancer. *Immun Ageing*. **15**, 1.

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. **12**, 323.

Li, C. H., Haider, S., & Boutros, P. C. (2022). Age influences on the molecular presentation of tumours. *Nat Commun*. **13**, 208.

Li, C. H., Haider, S., Shiah, Y. J., Thai, K., & Boutros, P. C. (2018). Sex Differences in Cancer Driver Genes and Biomarkers. *Cancer Res*. **78**, 5527-5537.

Li, C. M., Shapiro, H., Tsiobikas, C., Selfors, L. M., Chen, H., Rosenbluth, J., ... Brugge, J. S. (2020a). Aging-Associated Alterations in Mammary Epithelia and Stroma Revealed by Single-Cell RNA Sequencing. *Cell Rep*. **33**, 108566.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078-2079.

Li, R., Di, L., Li, J., Fan, W., Liu, Y., Guo, W., ... Wu, C. (2021a). A body map of somatic mutagenesis in morphologically normal human tissues. *Nature*. **597**, 398-403.

Li, R., Du, Y., Chen, Z., Xu, D., Lin, T., Jin, S., ... Bai, F. (2020b). Macroscopic somatic clonal expansion in morphologically normal human urothelium. *Science*. **370**, 82-89.

Li, Y., Roberts, N. D., Wala, J. A., Shapira, O., Schumacher, S. E., Kumar, K., ... Consortium, P. (2020c). Patterns of somatic structural variation in human cancer genomes. *Nature*. **578**, 112-121.

Li, Z., Zhang, Z., Ren, Y., Wang, Y., Fang, J., Yue, H., ... Guan, F. (2021b). Aging and age-related diseases: from mechanisms to therapeutic strategies. *Biogerontology*. **22**, 165-187.

Liao, S., Hartmaier, R. J., McGuire, K. P., Puhalla, S. L., Luthra, S., Chandran, U. R., ... Oesterreich, S. (2015). The molecular landscape of premenopausal breast cancer. *Breast Cancer Res*. **17**, 104.

Lieu, C. H., Golemis, E. A., Serebriiskii, I. G., Newberg, J., Hemmerich, A., Connelly, C., ... Meyer, J. E. (2019). Comprehensive Genomic Landscapes in Early and Later Onset Colorectal Cancer. *Clin Cancer Res*. **25**, 5852-5858.

Lipinski, K. A., Barber, L. J., Davies, M. N., Ashenden, M., Sottoriva, A., & Gerlinger, M. (2016). Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends Cancer*. **2**, 49-63.

Liu, D., & Hornsby, P. J. (2007). Senescent human fibroblasts increase the early growth of xenograft tumors via matrix metalloproteinase secretion. *Cancer Res*. **67**, 3117-3126.

Liu, J., Qu, S., Zhang, T., Gao, Y., Shi, H., Song, K., ... Yin, W. (2021). Applications of Single-Cell Omics in Tumor Immunology. *Front Immunol*. **12**, 697412.

Liu, L., van Groen, T., Kadish, I., Li, Y., Wang, D., James, S. R., ... Tollefsbol, T. O. (2011). Insufficient DNA methylation affects healthy aging and promotes age-related health problems. *Clin Epigenetics*. **2**, 349-360.

Liu, X., Lei, Q., Yu, Z., Xu, G., Tang, H., Wang, W., ... Wu, M. (2015). MiR-101 reverses the hypomethylation of the LMO3 promoter in glioma cells. *Oncotarget*. **6**, 7930-7943.

Liu, Y., Ao, X., Ding, W., Ponnusamy, M., Wu, W., Hao, X., ... Wang, J. (2018a). Critical role of FOXO3a in carcinogenesis. *Mol Cancer*. **17**, 104.

Liu, Y., Sethi, N. S., Hinoue, T., Schneider, B. G., Cherniack, A. D., Sanchez-Vega, F., ... Laird, P. W. (2018b). Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. *Cancer Cell*. **33**, 721-735 e728.

Lopez, S., Lim, E. L., Horswell, S., Haase, K., Huebner, A., Dietzen, M., ... McGranahan, N. (2020). Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution. *Nat Genet*. **52**, 283-293.

Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*. **153**, 1194-1217.

Lu, Y. X., Regan, J. C., Esser, J., Drews, L. F., Weinseis, T., Stinn, J., ... Partridge, L. (2021). A TORC1-histone axis regulates chromatin organisation and non-canonical induction of autophagy to ameliorate ageing. *Elife*. **10**.

Lui, J. C., Chen, W., Barnes, K. M., & Baron, J. (2010). Changes in gene expression associated with aging commonly originate during juvenile growth. *Mech Ageing Dev*. **131**, 641-649.

Ma, X., Liu, Y., Liu, Y., Alexandrov, L. B., Edmonson, M. N., Gawad, C., ... Zhang, J. (2018). Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*. **555**, 371-376.

Maas, H. A., Kruitwagen, R. F., Lemmens, V. E., Goey, S. H., & Janssen-Heijnen, M. L. (2005). The influence of age and co-morbidity on treatment and prognosis of ovarian cancer: a population-based study. *Gynecol Oncol*. **97**, 104-109.

Mabbott, N. A., Kobayashi, A., Sehgal, A., Bradford, B. M., Pattison, M., & Donaldson, D. S. (2015). Aging and the mucosal immune system in the intestine. *Biogerontology*. **16**, 133-145.

Macconaill, L. E., & Garraway, L. A. (2010). Clinical implications of the cancer genome. *J Clin Oncol*. **28**, 5219-5228.

Malaquin, N., Vercamer, C., Bouali, F., Martien, S., Deruy, E., Wernert, N., ... Pourtier, A. (2013). Senescent fibroblasts enhance early skin carcinogenic events via a paracrine MMP-PAR-1 axis. *PLoS One*. **8**, e63607.

Maria Chiara, B., Laura Di, R., Luca, G., Ilaria, C., Arianna, V., Margherita, R., ... Marco, B. (2021). Increased expression of Ectodysplasin A2 Receptor EDA2R is the most remarkable and ubiquitous aging-related transcriptional hallmark. *Nature Portfolio*.

Martincorena, I. (2019). Somatic mutation and clonal expansions in human tissues. *Genome Med*. **11**, 35.

Martincorena, I., & Campbell, P. J. (2015). Somatic mutation in cancer and normal cells. *Science*. **349**, 1483-1489.

Martincorena, I., Fowler, J. C., Wabik, A., Lawson, A. R. J., Abascal, F., Hall, M. W. J., ... Jones, P. H. (2018). Somatic mutant clones colonize the human esophagus with age. *Science*. **362**, 911-917.

Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., ... Campbell, P. J. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell*. **171**, 1029-1041 e1021.

Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., ... Campbell, P. J. (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*. **348**, 880-886.

Martinez-Jimenez, F., Muinos, F., Sentis, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., ... Lopez-Bigas, N. (2020). A compendium of mutational cancer driver genes. *Nat Rev Cancer*. **20**, 555-572.

Marttila, S., Kananen, L., Hayrynen, S., Jylhava, J., Nevalainen, T., Hervonen, A., ... Hurme, M. (2015). Ageing-associated changes in the human DNA methylome: genomic locations and effects on gene expression. *BMC Genomics*. **16**, 179.

Marusyk, A., & DeGregori, J. (2008). Declining cellular fitness with age promotes cancer initiation by selecting for adaptive oncogenic mutations. *Biochim Biophys Acta*. **1785**, 1-11.

Mayakonda, A., Lin, D. C., Assenov, Y., Plass, C., & Koeffler, H. P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. **28**, 1747-1756.

McCauley, B. S., & Dang, W. (2014). Histone methylation and aging: lessons learned from model systems. *Biochim Biophys Acta*. **1839**, 1454-1462.

McGranahan, N., & Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*. **168**, 613-628.

McKerrell, T., Park, N., Moreno, T., Grove, C. S., Ponstingl, H., Stephens, J., ... Vassiliou, G. S. (2015). Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep*. **10**, 1239-1245.

Mealey, N. E., O'Sullivan, D. E., Pader, J., Ruan, Y., Wang, E., Quan, M. L., & Brenner, D. R. (2020). Mutational landscape differences between young-onset and older-onset breast cancer patients. *BMC Cancer*. **20**, 212.

Medeiros Tavares Marques, J. C., Cornelio, D. A., Nogueira Silbiger, V., Ducati Luchessi, A., de Souza, S., & Batistuzzo de Medeiros, S. R. (2017). Identification of new genes associated to senescent and tumorigenic phenotypes in mesenchymal stem cells. *Sci Rep*. **7**, 17837.

Mehra, R., Varambally, S., Ding, L., Shen, R., Sabel, M. S., Ghosh, D., ... Kleer, C. G. (2005). Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res*. **65**, 11259-11264.

Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., & Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. **12**, R41.

Meynet, O., & Ricci, J. E. (2014). Caloric restriction and cancer: molecular mechanisms and clinical implications. *Trends Mol Med*. **20**, 419-427.
Milanovic, M., Fan, D. N. Y., Belenki, D., Dabritz, J. H. M., Zhao, Z., Yu, Y., ... Schmitt, C. A. (2018). Senescence-associated reprogramming promotes cancer stemness. *Nature*. **553**, 96-100.
Milholland, B., Auton, A., Suh, Y., & Vijg, J. (2015). Age-related somatic mutations in the cancer genome. *Oncotarget*. **6**, 24627-24635.
Mirchia, K., & Richardson, T. E. (2020). Beyond IDH-Mutation: Emerging Molecular Diagnostic and Prognostic Features in Adult Diffuse Gliomas. *Cancers (Basel)*. **12**.
Modrek, A. S., Golub, D., Khan, T., Bready, D., Prado, J., Bowman, C., ... Placantonakis, D. G. (2017). Low-Grade Astrocytoma Mutations in IDH1, P53, and ATRX Cooperate to Block Differentiation of Human Neural Stem Cells via Repression of SOX2. *Cell Rep*. **21**, 1267-1280.
Mogilenko, D. A., Shchukina, I., & Artyomov, M. N. (2021). Immune ageing at single-cell resolution. *Nat Rev Immunol*.
Moore, L., Cagan, A., Coorens, T. H. H., Neville, M. D. C., Sanghvi, R., Sanders, M. A., ... Rahbari, R. (2021). The mutational landscape of human somatic and germline cells. *Nature*. **597**, 381-386.
Moore, L., Leongamornlert, D., Coorens, T. H. H., Sanders, M. A., Ellis, P., Dentro, S. C., ... Stratton, M. R. (2020). The mutational landscape of normal human endometrial epithelium. *Nature*. **580**, 640-646.
Mudunuri, U., Che, A., Yi, M., & Stephens, R. M. (2009). bioDBnet: the biological database network. *Bioinformatics*. **25**, 555-556.
Munoz-Espin, D., Canamero, M., Maraver, A., Gomez-Lopez, G., Contreras, J., Murillo-Cuesta, S., ... Serrano, M. (2013). Programmed cell senescence during mammalian embryonic development. *Cell*. **155**, 1104-1118.
Nebbioso, A., Tambaro, F. P., Dell'Aversana, C., & Altucci, L. (2018). Cancer epigenetics: Moving forward. *PLoS Genet*. **14**, e1007362.
Neff, F., Flores-Dominguez, D., Ryan, D. P., Horsch, M., Schroder, S., Adler, T., ... Ehninger, D. (2013). Rapamycin extends murine lifespan but has limited effects on aging. *J Clin Invest*. **123**, 3272-3291.
Nelson, G., Wordsworth, J., Wang, C., Jurk, D., Lawless, C., Martin-Ruiz, C., & von Zglinicki, T. (2012). A senescent cell bystander effect: senescence-induced senescence. *Aging Cell*. **11**, 345-349.
Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science*. **194**, 23-28.
Oh, J., Lee, Y. D., & Wagers, A. J. (2014). Stem cell aging: mechanisms, regulators and therapeutic opportunities. *Nat Med*. **20**, 870-880.
Ordys, B. B., Launay, S., Deighton, R. F., McCulloch, J., & Whittle, I. R. (2010). The role of mitochondria in glioma pathophysiology. *Mol Neurobiol*. **42**, 64-75.
Osako, T., Lee, H., Turashvili, G., Chiu, D., McKinney, S., Joosten, S. E. P., ... Aparicio, S. (2020). Age-correlated protein and transcript expression in breast cancer and normal breast tissues is dominated by host endocrine effects. *Nature Cancer*. **1**, 518-532.
Oubaha, M., Miloudi, K., Dejda, A., Guber, V., Mawambo, G., Germain, M. A., ... Sapieha, P. (2016). Senescence-associated secretory phenotype contributes to pathological angiogenesis in retinopathy. *Sci Transl Med*. **8**, 362ra144.
Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., ... Tyers, M. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci*. **30**, 187-200.
Painter, R. C., Roseboom, T. J., & Bleker, O. P. (2005). Prenatal exposure to the Dutch famine and disease in later life: an overview. *Reprod Toxicol*. **20**, 345-352.

Pal, S., & Tyler, J. K. (2016). Epigenetics and aging. *Sci Adv*. **2**, e1600584.
Palmer, D., Fabris, F., Doherty, A., Freitas, A. A., & de Magalhaes, J. P. (2021). Ageing transcriptome meta-analysis reveals similarities and differences between key mammalian tissues. *Aging (Albany NY)*. **13**, 3313-3341.
Palmer, S., Albergante, L., Blackburn, C. C., & Newman, T. J. (2018). Thymic involution and rising disease incidence with age. *Proc Natl Acad Sci U S A*. **115**, 1883-1888.
Papadopoli, D., Boulay, K., Kazak, L., Pollak, M., Mallette, F., Topisirovic, I., & Hulea, L. (2019). mTOR as a central regulator of lifespan and aging. *F1000Res*. **8**.
Parikh, N., Shuck, R. L., Gagea, M., Shen, L., & Donehower, L. A. (2018). Enhanced inflammation and attenuated tumor suppressor pathways are associated with oncogene-induced lung tumors in aged mice. *Aging Cell*. **17**.
Partridge, L., Fuentealba, M., & Kennedy, B. K. (2020). The quest to slow ageing through drug discovery. *Nat Rev Drug Discov*. **19**, 513-532.
Pawelec, G. (2017). Immunosenescence and cancer. *Biogerontology*. **18**, 717-721.
Pazolli, E., Luo, X., Brehm, S., Carbery, K., Chung, J. J., Prior, J. L., ... Stewart, S. A. (2009). Senescent stromal-derived osteopontin promotes preneoplastic cell growth. *Cancer Res*. **69**, 1230-1239.
Perez, R. F., Tejedor, J. R., Bayon, G. F., Fernandez, A. F., & Fraga, M. F. (2018). Distinct chromatin signatures of DNA hypomethylation in aging and cancer. *Aging Cell*. **17**, e12744.
Podlaha, O., Riester, M., De, S., & Michor, F. (2012). Evolution of the cancer genome. *Trends Genet*. **28**, 155-163.
Power, C., & Jefferis, B. J. (2002). Fetal environment and subsequent obesity: a study of maternal smoking. *Int J Epidemiol*. **31**, 413-419.
Preston, J. D., Reynolds, L. J., & Pearson, K. J. (2018). Developmental Origins of Health Span and Life Span: A Mini-Review. *Gerontology*. **64**, 237-245.
Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841-842.
Raj, K., & Horvath, S. (2020). Current perspectives on the cellular and molecular features of epigenetic ageing. *Exp Biol Med (Maywood)*. **245**, 1532-1542.
Rakyan, V. K., Down, T. A., Maslau, S., Andrew, T., Yang, T. P., Beyan, H., ... Spector, T. D. (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res*. **20**, 434-439.
Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., ... Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. **44**, W160-165.
Rayfield, S., & Plugge, E. (2017). Systematic review and meta-analysis of the association between maternal smoking in pregnancy and childhood overweight and obesity. *J Epidemiol Community Health*. **71**, 162-173.
Rayon, T., & Briscoe, J. (2021). Cross-species comparisons and in vitro models to study tempo in development and homeostasis. *Interface Focus*. **11**, 20200069.
Ren, J., Stroncek, D. F., Zhao, Y., Jin, P., Castiello, L., Civini, S., ... Sabatino, M. (2013). Intra-subject variability in human bone marrow stromal cell (BMSC) replicative senescence: molecular changes associated with BMSC senescence. *Stem Cell Res*. **11**, 1060-1073.
Reynolds, L. M., Taylor, J. R., Ding, J., Lohman, K., Johnson, C., Siscovick, D., ... Liu, Y. (2014). Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nat Commun*. **5**, 5366.
Ribatti, D. (2017). The concept of immune surveillance against tumors. The first theories. *Oncotarget*. **8**, 7175-7180.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. **43**, e47.

Robertson, N. A., Latorre-Crespo, E., Terradas-Terradas, M., Purcell, A. C., Livesey, B. J., Marsh, J. A., ... Chandra, T. (2021). Longitudinal dynamics of clonal hematopoiesis identifies gene-specific fitness effects. *bioRxiv*, 2021.2005.2027.446006.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139-140.

Robinson, P. S., Coorens, T. H. H., Palles, C., Mitchell, E., Abascal, F., Olafsson, S., ... Stratton, M. R. (2021a). Increased somatic mutation burdens in normal human cells due to defective DNA polymerases. *Nat Genet*. **53**, 1434-1442.

Robinson, P. S., Thomas, L. E., Abascal, F., Jung, H., Harvey, L. M. R., West, H. D., ... Stratton, M. R. (2021b). Inherited MUTYH mutations cause elevated somatic mutation rates and distinctive mutational signatures in normal human cells. *bioRxiv*, 2021.2010.2020.465093.

Rojas, A., Liu, G., Coleman, I., Nelson, P. S., Zhang, M., Dash, R., ... Wu, J. D. (2011). IL-6 promotes prostate tumorigenesis and progression through autocrine cross-activation of IGF-IR. *Oncogene*. **30**, 2345-2355.

Rose, M. R., Flatt, T., Graves, J. L., Greer, L. F., Martinez, D. E., Matos, M., ... Shahrestani, P. (2012). What is Aging? *Front Genet*. **3**, 134.

Rossi, D. J., Jamieson, C. H., & Weissman, I. L. (2008). Stems cells and the pathways to aging and cancer. *Cell*. **132**, 681-696.

Rossiello, F., Jurk, D., Passos, J. F., & d'Adda di Fagagna, F. (2022). Telomere dysfunction in ageing and age-related diseases. *Nat Cell Biol*. **24**, 135-147.

Rozhok, A., & DeGregori, J. (2019). A generalized theory of age-dependent carcinogenesis. *Elife*. **8**.

Rozhok, A. I., Salstrom, J. L., & DeGregori, J. (2014). Stochastic modeling indicates that aging and somatic evolution in the hematopoetic system are driven by non-cell-autonomous processes. *Aging (Albany NY)*. **6**, 1033-1048.

Ryland, G. L., Doyle, M. A., Goode, D., Boyle, S. E., Choong, D. Y., Rowley, S. M., ... Gorringe, K. L. (2015). Loss of heterozygosity: what is it good for? *BMC Med Genomics*. **8**, 45.

Sacher, A. G., Dahlberg, S. E., Heng, J., Mach, S., Janne, P. A., & Oxnard, G. R. (2016). Association Between Younger Age and Targetable Genomic Alterations and Prognosis in Non-Small-Cell Lung Cancer. *JAMA Oncol*. **2**, 313-320.

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W. K., Luna, A., La, K. C., ... Schultz, N. (2018). Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell*. **173**, 321-337 e310.

Santarosa, M., Del Col, L., Tonin, E., Caragnano, A., Viel, A., & Maestro, R. (2009). Premature senescence is a major response to DNA cross-linking agents in BRCA1-defective cells: implication for tailored treatments of BRCA1 mutation carriers. *Mol Cancer Ther*. **8**, 844-854.

Santos, J., & Gil, J. (2014). TRIM28/KAP1 regulates senescence. *Immunol Lett*. **162**, 281-289.

Schneider, G., Schmidt-Supprian, M., Rad, R., & Saur, D. (2017). Tissue-specific tumorigenesis: context matters. *Nat Rev Cancer*. **17**, 239-253.

Schwarze, S. R., DePrimo, S. E., Grabert, L. M., Fu, V. X., Brooks, J. D., & Jarrard, D. F. (2002). Novel pathways associated with bypassing cellular senescence in human prostate epithelial cells. *J Biol Chem*. **277**, 14877-14883.

Sen, P., Shah, P. P., Nativio, R., & Berger, S. L. (2016). Epigenetic Mechanisms of Longevity and Aging. *Cell*. **166**, 822-839.

Serrano, M., & Blasco, M. A. (2007). Cancer and ageing: convergent and divergent mechanisms. *Nat Rev Mol Cell Biol*. **8**, 715-722.

Shah, P. P., Donahue, G., Otte, G. L., Capell, B. C., Nelson, D. M., Cao, K., ... Berger, S. L. (2013). Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev*. **27**, 1787-1799.

Shah, Y., Verma, A., Marderstein, A. R., White, J., Bhinder, B., Garcia Medina, J. S., & Elemento, O. (2021). Pan-cancer analysis reveals molecular patterns associated with age. *Cell Rep*. **37**, 110100.

Shindyapina, A. V., Cho, Y., Kaya, A., Tyshkovskiy, A., Castro, J. P., Gordevicius, J., ... Gladyshev, V. N. (2022). Rapamycin treatment during development extends lifespan and healthspan. *bioRxiv*, 2022.2002.2018.481092.

Shlien, A., Campbell, B. B., de Borja, R., Alexandrov, L. B., Merico, D., Wedge, D., ... Biallelic Mismatch Repair Deficiency, C. (2015). Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. *Nat Genet*. **47**, 257-262.

Silva, A. S., Wood, S. H., van Dam, S., Berres, S., McArdle, A., & de Magalhaes, J. P. (2011). Gathering insights on disease etiology from gene expression profiles of healthy tissues. *Bioinformatics*. **27**, 3300-3305.

Silver, A. J., & Jaiswal, S. (2019). Clonal hematopoiesis: Pre-cancer PLUS. *Adv Cancer Res*. **141**, 85-128.

Simpkin, A. J., Hemani, G., Suderman, M., Gaunt, T. R., Lyttleton, O., McArdle, W. L., ... Smith, G. D. (2016). Prenatal and early life influences on epigenetic age in children: a study of mother-offspring pairs from two cohort studies. *Hum Mol Genet*. **25**, 191-201.

Simpson, C. L., & Al-Chalabi, A. (2006). Amyotrophic lateral sclerosis as a complex genetic disease. *Biochim Biophys Acta*. **1762**, 973-985.

Sinha, S., Mitchell, K. A., Zingone, A., Bowman, E., Sinha, N., Schäffer, A. A., ... Ryan, B. M. (2020). Higher prevalence of homologous recombination deficiency in tumors from African Americans versus European Americans. *Nature Cancer*. **1**, 112-121.

Slowikowski, K. (2020). ggrepel: Automatically Position Non-Overlapping Text Labels with 'ggplot2'.

Smith, J. C., & Sheltzer, J. M. (2018). Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. *Elife*. **7**.

So, K., Tamura, G., Honda, T., Homma, N., Waki, T., Togawa, N., ... Motoyama, T. (2006). Multiple tumor suppressor genes are increasingly methylated with age in non-neoplastic gastric epithelia. *Cancer Sci*. **97**, 1155-1158.

Somel, M., Guo, S., Fu, N., Yan, Z., Hu, H. Y., Xu, Y., ... Khaitovich, P. (2010). MicroRNA, mRNA, and protein expression link development and aging in human and macaque brain. *Genome Res*. **20**, 1207-1218.

Song, L., Rawal, B., Nemeth, J. A., & Haura, E. B. (2011). JAK1 activates STAT3 activity in non-small-cell lung cancer cells and IL-6 neutralizing antibodies can suppress JAK1-STAT3 signaling. *Mol Cancer Ther*. **10**, 481-494.

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat Rev Genet*. **20**, 631-656.

Steegenga, W. T., Boekschoten, M. V., Lute, C., Hooiveld, G. J., de Groot, P. J., Morris, T. J., ... Muller, M. (2014). Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age (Dordr)*. **36**, 9648.

Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*. **458**, 719-724.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. **102**, 15545-15550.

Suh, Y., Atzmon, G., Cho, M. O., Hwang, D., Liu, B., Leahy, D. J., ... Cohen, P. (2008). Functionally significant insulin-like growth factor I receptor mutations in centenarians. *Proc Natl Acad Sci U S A*. **105**, 3438-3442.

Sweeney, C., Bernard, P. S., Factor, R. E., Kwan, M. L., Habel, L. A., Quesenberry, C. P., Jr., ... Caan, B. J. (2014). Intrinsic subtypes from PAM50 gene expression assay in a population-based breast cancer cohort: differences by age, race, and tumor characteristics. *Cancer Epidemiol Biomarkers Prev*. **23**, 714-724.

Tabula Muris, C. (2020). A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*. **583**, 590-595.

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., ... de Magalhaes, J. P. (2018). Human Ageing Genomic Resources: new and updated databases. *Nucleic Acids Res*. **46**, D1083-D1090.

Takahashi, A., Imai, Y., Yamakoshi, K., Kuninaka, S., Ohtani, N., Yoshimoto, S., ... Hara, E. (2012). DNA damage signaling triggers degradation of histone methyltransferases through APC/C(Cdh1) in senescent cells. *Mol Cell*. **45**, 123-131.

Takaku, M., Grimm, S. A., Roberts, J. D., Chrysovergis, K., Bennett, B. D., Myers, P., ... Wade, P. A. (2018). GATA3 zinc finger 2 mutations reprogram the breast cancer transcriptional network. *Nat Commun*. **9**, 1059.

Takaku, M., Grimm, S. A., & Wade, P. A. (2015). GATA3 in Breast Cancer: Tumor Suppressor or Oncogene? *Gene Expr*. **16**, 163-168.

Takasugi, M. (2018). Emerging roles of extracellular vesicles in cellular senescence and aging. *Aging Cell*. **17**.

Takasugi, M., Okada, R., Takahashi, A., Virya Chen, D., Watanabe, S., & Hara, E. (2017). Small extracellular vesicles secreted from senescent cells promote cancer cell proliferation through EphA2. *Nat Commun*. **8**, 15729.

Tanda, E. T., Vanni, I., Boutros, A., Andreotti, V., Bruno, W., Ghiorzo, P., & Spagnolo, F. (2020). Current State of Target Treatment in BRAF Mutated Melanoma. *Front Mol Biosci*. **7**, 154.

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., ... Forbes, S. A. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res*. **47**, D941-D947.

Team, R. C. (2020). R: A Language and Environment for Statistical Computing.

Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Weisenberger, D. J., Shen, H., ... Widschwendter, M. (2010). Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res*. **20**, 440-446.

Tomas-Loba, A., Flores, I., Fernandez-Marcos, P. J., Cayuela, M. L., Maraver, A., Tejera, A., ... Blasco, M. A. (2008). Telomerase reverse transcriptase delays aging in cancer-resistant mice. *Cell*. **135**, 609-622.

Tomasetti, C., Poling, J., Roberts, N. J., London, N. R., Jr., Pittman, M. E., Haffner, M. C., ... Vogelstein, B. (2019). Cell division rates decrease with age, providing a potential explanation for the age-dependent deceleration in cancer incidence. *Proc Natl Acad Sci U S A*. **116**, 20482-20488.

Tomasetti, C., Vogelstein, B., & Parmigiani, G. (2013). Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc Natl Acad Sci U S A*. **110**, 1999-2004.

Toso, A., Revandkar, A., Di Mitri, D., Guccini, I., Proietti, M., Sarti, M., ... Alimonti, A. (2014). Enhancing chemotherapy efficacy in Pten-deficient prostate tumors by activating the senescence-associated antitumor immunity. *Cell Rep*. **9**, 75-89.

Tuttle, C. S. L., Luesken, S. W. M., Waaijer, M. E. C., & Maier, A. B. (2021). Senescence in tissue samples of humans with age-related diseases: A systematic review. *Ageing Res Rev*. **68**, 101334.

Tuttle, C. S. L., Waaijer, M. E. C., Slee-Valentijn, M. S., Stijnen, T., Westendorp, R., & Maier, A. B. (2020). Cellular senescence and chronological age in various human tissues: A systematic review and meta-analysis. *Aging Cell*. **19**, e13083.

Valdes, A. M., Glass, D., & Spector, T. D. (2013). Omics technologies and the study of human ageing. *Nat Rev Genet*. **14**, 601-607.

Van de Peer, Y., Mizrachi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. *Nat Rev Genet*. **18**, 411-424.

van Deursen, J. M. (2014). The role of senescent cells in ageing. *Nature*. **509**, 439-446.

Van Herck, Y., Feyaerts, A., Alibhai, S., Papamichael, D., Decoster, L., Lambrechts, Y., ... Wildiers, H. (2021). Is cancer biology different in older patients? *The Lancet Healthy Longevity*. **2**, e663-e677.

Van Loo, P., Nordgard, S. H., Lingjaerde, O. C., Russnes, H. G., Rye, I. H., Sun, W., ... Kristensen, V. N. (2010). Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A*. **107**, 16910-16915.

Vas, V., Wandhoff, C., Dorr, K., Niebel, A., & Geiger, H. (2012). Contribution of an aged microenvironment to aging-associated myeloproliferative disease. *PLoS One*. **7**, e31523.

Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., ... Cancer Genome Atlas Research, N. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. **17**, 98-110.

Vijg, J., & Dong, X. (2020). Pathogenic Mechanisms of Somatic Mutation and Genome Mosaicism in Aging. *Cell*. **182**, 12-23.

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science*. **339**, 1546-1558.

Wadhwa, P. D., Buss, C., Entringer, S., & Swanson, J. M. (2009). Developmental origins of health and disease: brief history of the approach and current focus on epigenetic mechanisms. *Semin Reprod Med*. **27**, 358-368.

Waks, A. G., Kim, D., Jain, E., Snow, C., Kirkner, G. J., Rosenberg, S. M., ... Wagle, N. (2022). Somatic and Germline Genomic Alterations in Very Young Women with Breast Cancer. *Clin Cancer Res*. **28**, 2339-2348.

Wang, H., Wang, X., Xu, L., Zhang, J., & Cao, H. (2020). Prognostic significance of age related genes in patients with lower grade glioma. *J Cancer*. **11**, 3986-3999.

Wang, Z., Chivu, A. G., Choate, L. A., Rice, E. J., Miller, D. C., Chu, T., ... Danko, C. G. (2022). Prediction of histone post-translational modification patterns based on nascent transcription data. *Nat Genet*. **54**, 295-305.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. **10**, 57-63.

Watson, C. J., Papula, A. L., Poon, G. Y. P., Wong, W. H., Young, A. L., Druley, T. E., ... Blundell, J. R. (2020). The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science*. **367**, 1449-1454.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York.

Wijewardhane, N., Dressler, L., & Ciccarelli, F. D. (2021). Normal Somatic Mutations in Cancer Transformation. *Cancer Cell*. **39**, 125-129.
Wiley, C. D., Flynn, J. M., Morrissey, C., Lebofsky, R., Shuga, J., Dong, X., ... Campisi, J. (2017). Analysis of individual cells identifies cell-to-cell variability following induction of cellular senescence. *Aging Cell*. **16**, 1043-1050.
Wiley, C. D., Schaum, N., Alimirah, F., Lopez-Dominguez, J. A., Orjalo, A. V., Scott, G., ... Campisi, J. (2018). Small-molecule MDM2 antagonists attenuate the senescence-associated secretory phenotype. *Sci Rep*. **8**, 2410.
Wilkinson, J. E., Burmeister, L., Brooks, S. V., Chan, C. C., Friedline, S., Harrison, D. E., ... Miller, R. A. (2012). Rapamycin slows aging in mice. *Aging Cell*. **11**, 675-682.
Willauer, A. N., Liu, Y., Pereira, A. A. L., Lam, M., Morris, J. S., Raghav, K. P. S., ... Loree, J. M. (2019). Clinical and molecular characterization of early-onset colorectal cancer. *Cancer*. **125**, 2002-2010.
Willcox, B. J., Donlon, T. A., He, Q., Chen, R., Grove, J. S., Yano, K., ... Curb, J. D. (2008). FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci U S A*. **105**, 13987-13992.
Williams, G. C. (1957). Pleiotropy, Natural Selection, and the Evolution of Senescence. *Evolution*. **11**, 398-411.
Wong, K. M., Hudson, T. J., & McPherson, J. D. (2011). Unraveling the genetics of cancer: genome sequencing and beyond. *Annu Rev Genomics Hum Genet*. **12**, 407-430.
Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., ... Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (N Y)*. **2**, 100141.
Wu, Y., Wei, J., Chen, X., Qin, Y., Mao, R., Song, J., & Fan, Y. (2019). Comprehensive transcriptome profiling in elderly cancer patients reveals aging-altered immune cells and immune checkpoints. *Int J Cancer*. **144**, 1657-1663.
Xie, M., Lu, C., Wang, J., McLellan, M. D., Johnson, K. J., Wendl, M. C., ... Ding, L. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med*. **20**, 1472-1478.
Xu, F., Sun, S., Yan, S., Guo, H., Dai, M., & Teng, Y. (2015). Elevated expression of RIT1 correlates with poor prognosis in endometrial cancer. *Int J Clin Exp Pathol*. **8**, 10315-10324.
Xu, M., Pirtskhalava, T., Farr, J. N., Weigand, B. M., Palmer, A. K., Weivoda, M. M., ... Kirkland, J. L. (2018). Senolytics improve physical function and increase lifespan in old age. *Nat Med*. **24**, 1246-1256.
Yan, H., Parsons, D. W., Jin, G., McLendon, R., Rasheed, B. A., Yuan, W., ... Bigner, D. D. (2009). IDH1 and IDH2 mutations in gliomas. *N Engl J Med*. **360**, 765-773.
Yang, H., Wang, H., Ren, J., Chen, Q., & Chen, Z. J. (2017). cGAS is essential for cellular senescence. *Proc Natl Acad Sci U S A*. **114**, E4612-E4620.
Yang, H., Zhong, Y., Peng, C., Chen, J. Q., & Tian, D. (2010). Important role of indels in somatic mutations of human cancer genes. *BMC Med Genet*. **11**, 128.
Yang, J., Huang, T., Petralia, F., Long, Q., Zhang, B., Argmann, C., ... Consortium, G. T. (2015). Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci Rep*. **5**, 15145.
Yau, C., Fedele, V., Roydasgupta, R., Fridlyand, J., Hubbard, A., Gray, J. W., ... Benz, C. C. (2007). Aging impacts transcriptomes but not genomes of hormone-dependent breast cancers. *Breast Cancer Res*. **9**, R59.
Yi, S. J., & Kim, K. (2020). New Insights into the Role of Histone Changes in Aging. *Int J Mol Sci*. **21**.

Yizhak, K., Aguet, F., Kim, J., Hess, J. M., Kubler, K., Grimsby, J., ... Getz, G. (2019). RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science*. **364**.

Yokoyama, A., Kakiuchi, N., Yoshizato, T., Nannya, Y., Suzuki, H., Takeuchi, Y., ... Ogawa, S. (2019). Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature*. **565**, 312-317.

Yoshida, K., Gowers, K. H. C., Lee-Six, H., Chandrasekharan, D. P., Coorens, T., Maughan, E. F., ... Campbell, P. J. (2020). Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature*. **578**, 266-272.

Yousefzadeh, M. J., Zhao, J., Bukata, C., Wade, E. A., McGowan, S. J., Angelini, L. A., ... Niedernhofer, L. J. (2020). Tissue specificity of senescent cell accumulation during physiologic and accelerated aging of mice. *Aging Cell*. **19**, e13094.

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. **16**, 284-287.

Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. **31**, 2382-2383.

Yuan, J., Hu, Z., Mahal, B. A., Zhao, S. D., Kensler, K. H., Pi, J., ... Zhang, L. (2018). Integrated Analysis of Genetic Ancestry and Genomic Alterations across Cancers. *Cancer Cell*. **34**, 549-560 e549.

Yuan, Y., Liu, L., Chen, H., Wang, Y., Xu, Y., Mao, H., ... Liang, H. (2016). Comprehensive Characterization of Molecular Differences in Cancer between Male and Female Patients. *Cancer Cell*. **29**, 711-722.

Zabransky, D. J., Jaffee, E. M., & Weeraratna, A. T. (2022). Shared genetic and epigenetic changes link aging and cancer. *Trends Cell Biol*. **32**, 338-350.

Zack, T. I., Schumacher, S. E., Carter, S. L., Cherniack, A. D., Saksena, G., Tabak, B., ... Beroukhim, R. (2013). Pan-cancer patterns of somatic copy number alteration. *Nat Genet*. **45**, 1134-1140.

Zane, L., Sharma, V., & Misteli, T. (2014). Common features of chromatin in aging and cancer: cause or coincidence? *Trends Cell Biol*. **24**, 686-694.

Zentner, G. E., & Henikoff, S. (2014). High-resolution digital profiling of the epigenome. *Nat Rev Genet*. **15**, 814-827.

Zhang, H., Pan, K. H., & Cohen, S. N. (2003). Senescence-specific gene expression fingerprints reveal cell-type-dependent physical clustering of up-regulated chromosomal loci. *Proc Natl Acad Sci U S A*. **100**, 3251-3256.

Zhang, L., Dong, X., Lee, M., Maslov, A. Y., Wang, T., & Vijg, J. (2019). Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc Natl Acad Sci U S A*. **116**, 9014-9019.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., ... Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. **9**, R137.

Zhao, S., Allis, C. D., & Wang, G. G. (2021). The language of chromatin modification in human cancers. *Nat Rev Cancer*. **21**, 413-430.

Zhao, Z., & Shilatifard, A. (2019). Epigenetic modifications of histones in cancer. *Genome Biol*. **20**, 245.

Zink, F., Stacey, S. N., Norddahl, G. L., Frigge, M. L., Magnusson, O. T., Jonsdottir, I., ... Stefansson, K. (2017). Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood*. **130**, 742-752.

Zou, Z., Tao, T., Li, H., & Zhu, X. (2020). mTOR signaling pathway and mTOR inhibitors in cancer: progress and challenges. *Cell Biosci*. **10**, 31.